

Université de Montréal

Probabilité de fixation dans des modèles de
sélection avec consanguinité

par

Samuel Langevin

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en mathématiques

décembre 2006



AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

**Probabilité de fixation dans des modèles de
sélection avec consanguinité**

présenté par

Samuel Langevin

a été évalué par un jury composé des personnes suivantes :

Martin Goldstein

(président-rapporteur)

Sabin Lessard

(directeur de recherche)

Alcjandro Murua

(membre du jury)

Mémoire accepté le:

12 décembre 2006

SOMMAIRE

La probabilité de fixation d'un nouveau gène mutant pour un modèle diploïde simple de Wright-Fisher avec sélection est approximée par un processus de diffusion. Cette approximation est comparée, lorsque la taille de la population est grande, à la probabilité de fixation obtenue par une méthode exacte en utilisant la théorie de la coalescence. Ensuite, le calcul de la probabilité de fixation dans un modèle de viabilité pour une population diploïde de Wright-Fisher avec autofécondation partielle et interaction entre les individus est obtenu avec une autre méthode qui, elle, utilise le lemme de Möhle en plus de la théorie de la coalescence. Cette probabilité est utilisée pour obtenir les stratégies stables au cours de l'évolution (ESS) dans cette population. On fait de même pour un modèle diploïde de fertilité.

Mots-clefs : Probabilité de fixation, Stratégies stables au cours de l'évolution, Modèle de Wright-Fisher, Modèle d'autofécondation partielle, Théorie de la coalescence, Approximation par un processus de diffusion, Lemme de Möhle, Théorie dynamique des jeux.

SUMMARY

The fixation probability of a single new mutant is obtained by a diffusion approximation for a simple diploid population following the Wright-Fisher model with selection. This approximation is compared, when the population is large but finite, to the probability found by an exact method that uses coalescent theory. Then, the fixation probability in a diploid viability model of the Wright-Fisher type with partial selfing and interaction between individuals is obtained with a method that uses Möhle's lemma and coalescent theory. Evolutionarily stable strategies (ESS) are deduced from this probability. This is also done for a fertility model.

Keywords : Fixation probability, Evolutionarily stable strategies, Wright-Fisher model, Partial selfing model, Coalescent theory, Diffusion approximation, Möhle's lemma, Evolutionary game theory.

TABLE DES MATIÈRES

Sommaire	iii
Summary	iv
Liste des figures	vii
Remerciements	1
Introduction	2
Chapitre 1. Modèle neutre de Wright-Fisher	4
1.1. Description du modèle.....	4
1.2. Éléments de la théorie de la coalescence.....	5
Chapitre 2. Sélection et probabilité de fixation	13
2.1. Modèle de sélection.....	13
2.2. Approximation de la probabilité de fixation par un processus de diffusion.....	14
2.3. Approche directe pour la probabilité de fixation par la théorie de la coalescence.....	24
Chapitre 3. Probabilité de fixation dans un modèle de sélection avec interaction et consanguinité	41
3.1. Théorie dynamique des jeux.....	41
3.2. Probabilité de fixation dans une population avec autofécondation partielle.....	49

3.2.1. Probabilité de fixation dans une population avec autofécondation partielle et interaction	49
3.2.2. Analyse de la probabilité de fixation	89
3.2.3. Probabilité de fixation dans une population avec autofécondation partielle sans interaction	92
3.2.4. Probabilité de fixation dans un modèle de fertilité.....	95
Conclusion.....	100
Bibliographie	102
Annexe A. Calcul exact de l'espérance de T_3 lorsque $s = 0$.....	A-i
Annexe B. Les matrices	B-i
B.1. Matrice H dans le cas de l'autofécondation partielle	B-i
B.2. Matrice P dans le cas de l'autofécondation partielle.....	B-v
B.3. Matrice R_N dans le cas de l'autofécondation partielle	B-vii
B.4. Matrice R dans le cas de l'autofécondation partielle.....	B-x
B.5. Matrice $e^{PRP\tau}$ dans le cas de l'autofécondation partielle.....	B-xii
Annexe C. Lemme de Möhle et autofécondation partielle	C-i

LISTE DES FIGURES

1.1	Probabilités pour les gènes parentaux de deux gènes de la génération $t + 1$	6
1.2	Arbre de coalescence pour cinq gènes.....	10
3.1	La matrice de paiement.....	43
3.2	Les quinze états possibles pour la configuration ancestrale de six gènes.	60
3.3	Façon d'avoir le génotype AB à la génération t	61
3.4	Façon d'avoir la génotype AB au temps τ en nombre de $2N$ générations lorsque N tend vers l'infini.....	63
3.5	Façons d'avoir deux individus de génotypes AA et AB au temps τ en nombre de $2N$ générations lorsque N tend vers l'infini.....	65
3.6	Trois façons d'avoir deux individus de génotypes AA et BB au temps τ en nombre de $2N$ générations lorsque N tend vers l'infini.....	72
3.7	Trois autres façons d'avoir deux individus de génotypes AA et BB au temps τ en nombre de $2N$ générations lorsque N tend vers l'infini. ...	73
C.1	$\Pi^{[2N\tau]}$	C-iii
C.2	Les trois états possibles pour la configuration ancestrale de deux gènes	C-iii

REMERCIEMENTS

Je voudrais remercier M. Sabin Lessard, mon directeur de recherche, qui m'a fait découvrir ce domaine des mathématiques et qui a su m'encourager pour que je solutionne certains problèmes difficiles de cette recherche. Je remercie aussi Véronique Ladret pour ses conseils qui m'ont éclairé et qui m'ont permis de trouver quelques erreurs. Pour la contribution financière, je remercie encore M. Sabin Lessard, ainsi que mes parents, André Langevin et Sylvie Wilson, sans lesquels je serais encore au baccalauréat avec un cours par session.

Je tiens à remercier Véronique Sangin, pour son affection, et Catherine Andrewes, avec qui j'ai partagé les plus belles études universitaires. Je n'oublie pas Émilie Andrewes, Andrew Baum, Dominique Chayer, Pascale Chayer, Gérard Kubat, Maryse Langevin, Yanick Langevin, Susan Reitz et tout ceux qui m'ont fait décrocher des mathématiques, pour que j'y revienne plus fort. Aussi, pour leur assistance technique, je remercie Nicolas Beauchemin, Francis Forget et Alexandre Girouard.

INTRODUCTION

Charles Darwin (1809-1882) a fait prendre tout un tournant à la théorie de l'évolution avec la sélection naturelle ([D]). Avant celle-ci, c'était la théorie de Jean-Baptiste de Lamarck (1744-1829) qui était adoptée. Selon cette théorie, les caractéristiques génétiques des individus se perfectionnent par un usage répété et s'altèrent ou disparaissent si l'individu n'utilise pas ce caractère ([L]). Donc, d'une certaine manière, tous les individus peuvent "choisir" d'évoluer. Avec la théorie de Darwin, ce ne sont pas les individus qui sont les maîtres de leur évolution. Selon lui, les individus les mieux adaptés à leur environnement ont plus de chances de survivre et de se reproduire et donc, à long terme, les caractères qui restent dans la population sont ceux qui avantagent les individus. Ce sont donc les lois de la nature et les mutations qui seraient à la base de l'évolution.

De nos jours, la théorie de Darwin est, sans conteste, la théorie préférée des évolutionnistes. Cependant, après avoir accepté cette théorie, on peut se poser une question intéressante : Quelle est l'intensité de l'avantage sélectif qu'un nouveau gène mutant doit avoir pour rester dans la population et éliminer les autres gènes moins avantageux ? En fait ce n'est pas automatique que le nouveau gène mutant reste dans la population même s'il est beaucoup favorisé car comme il est nouveau, il n'est représenté qu'une seule fois dans toute la population. Il risque donc de disparaître et ce, très rapidement. Cela amène à une probabilité. Une probabilité, pour ce gène, de rester dans la population et d'éliminer ses compétiteurs : c'est la probabilité de fixation.

Beaucoup de travaux ont été faits à propos de cette probabilité. Il y a plusieurs façons de la trouver et aussi beaucoup de modèles de population auxquels on peut l'appliquer (voir [K], [E], [Le(2)] et [LL]). Jusqu'à présent, la probabilité de fixation a une utilité plutôt théorique car elle sert d'outil pour trouver des résultats plus intéressants. Pour une population de taille finie par exemple, on peut utiliser la probabilité de fixation pour trouver des stratégies stables au cours de l'évolution (voir [HS]). Ces dernières sont des stratégies, employées par les individus d'une population lors de rencontres ou de conflits, telles qu'il n'existe pas de stratégie meilleure pouvant les supplanter. Avec les stratégies stables au cours de l'évolution (ESS), on peut aussi comprendre la stabilité des états d'une population dynamique.

Dans ce mémoire, on s'intéressera aux méthodes pour trouver la probabilité de fixation et aux ESS. Au premier chapitre, on donnera une explication du modèle de population neutre de Wright-Fisher ([F], [W]) et de la théorie de la coalescence ([Ki(a)], [Ki(b)]), deux éléments fondamentaux de la génétique mathématique en population finie. Ensuite, au chapitre deux, on introduira la sélection et on verra l'approche approximative par le processus de diffusion (comme dans [Le(2)]) que l'on comparera à l'approche directe et exacte (par la théorie de la coalescence) de la probabilité de fixation ([LL]). Enfin, au chapitre trois, on passera par la théorie de la coalescence pour trouver la probabilité de fixation, qui, elle, nous aidera à trouver des stratégies stables au cours de l'évolution (voir [Le(2)]) dans le contexte de la théorie dynamique des jeux ([VM]). Dans ce dernier chapitre dont le travail est original, nous utiliserons deux modèles de Wright-Fisher avec autofécondation partielle : un dans lequel la sélection s'exprimera à travers des différences de viabilité et l'autre dans lequel elle s'exprimera à travers des différences de fertilité.

Chapitre 1

MODÈLE NEUTRE DE WRIGHT-FISHER

1.1. DESCRIPTION DU MODÈLE

Le modèle de Wright-Fisher ([F] en 1930 et [W] en 1931) est un des modèles les plus simples pour simuler l'évolution de la composition génique d'une population. Supposons une population diploïde (i.e. chaque individu a deux ensembles complets de gènes) de taille fixe N , donc de $2N$ gènes, des générations discrètes sans chevauchement, et absence de mutation d'un type à un autre et absence d'avantage sélectif d'un type sur un autre. On suppose aussi que la nouvelle génération remplace la précédente qui disparaît. Supposons qu'on a, au locus (la localisation du gène) considéré, seulement deux allèles A et B (types de gène ayant des effets différents). Il y a donc trois génotypes différents dans la population : AA , AB et BB . Soient $x = i/2N$ la fréquence de A à la génération t et $(1 - x) = (2N - i)/2N$ celle de B . Chaque individu diploïde transmet une copie de l'un de ses deux gènes du locus considéré à chaque gamète qu'il produit (les gamètes contiennent seulement un ensemble complet de gènes). De plus, chaque individu de la génération t produit un nombre infini de gamètes et les gamètes de tous les individus sont regroupées ensemble. Les fréquences de A et B restent donc les mêmes. Les gamètes sont ensuite réunis au hasard en paire. La fréquence des paires AA est donc x^2 , celle des paires AB est $2x(1 - x)$, et celle des paires BB est $(1 - x)^2$. On pige ensuite au hasard N paires de gamètes qui formeront la génération $t + 1$. Le nombre de gènes de type A à la génération $t + 1$ suit une

loi binomiale $(2N, x)$, puisque la probabilité de passer de i à j gènes de type A de la génération t à la génération $t + 1$ est

$$\binom{2N}{j} x^j (1-x)^{2N-j}, \quad 0 \leq j \leq 2N,$$

alors que les nombres d'individus de génotype AA, AB, BB , respectivement, suivent une loi multinomiale, i.e. que la probabilité qu'il y ait l individus de type AA , m de type AB , et n de type BB est

$$\binom{N}{l, m, n} (x^2)^l (2x(1-x))^m ((1-x)^2)^n,$$

où $l, m, n \geq 0$ et $l + m + n = N$.

1.2. ÉLÉMENTS DE LA THÉORIE DE LA COALESCENCE

La théorie de la coalescence est utilisée depuis une vingtaine d'années ([Ki(a)], [Ki(b)]) en génétique des populations. Elle permet d'observer les relations généalogiques d'un échantillon de gènes. Une coalescence entre deux gènes est une jonction des lignées de ceux-ci lorsqu'on remonte le temps. Voici un exemple pour bien l'illustrer. On prend deux individus humains (diploïdes), des soeurs par exemple, et on s'intéresse à un seul locus. Chaque fille a deux gènes à ce locus, un qui vient du père et un qui vient de la mère. Si les deux gènes provenant de la mère (un appartenant à la première fille et l'autre à la deuxième) sont issus du même gène maternel (car la mère a aussi deux gènes à ce locus et pour chaque enfant, elle n'en transmet qu'un), alors ils ont le même gène "ancêtre". En inversant l'échelle du temps, on observe que les deux gènes se fusionnent, on dit qu'il y a coalescence. En fait, dans la théorie de la coalescence, comme on veut

étudier les relations généalogiques d'un échantillon de gènes, on doit obligatoirement observer leur passé et donc remonter le temps.

Considérons un échantillon de $2N$ gènes suivant le modèle de Wright-Fisher sans mutation ni sélection. Supposons que la génération $t + 1$ est composée des enfants des individus de la génération t . En remontant le temps, la probabilité que deux gènes en particulier de la génération $t + 1$ coalescent à la génération t est $1/2N$ car, connaissant le gène parental du premier, la probabilité que le deuxième ait le même est $1/2N$ (voir la figure 1.1). On a donc que la probabilité que le nombre de générations jusqu'à la coalescence à partir de deux gènes soit plus grand que r est $(1 - \frac{1}{2N})^r$.

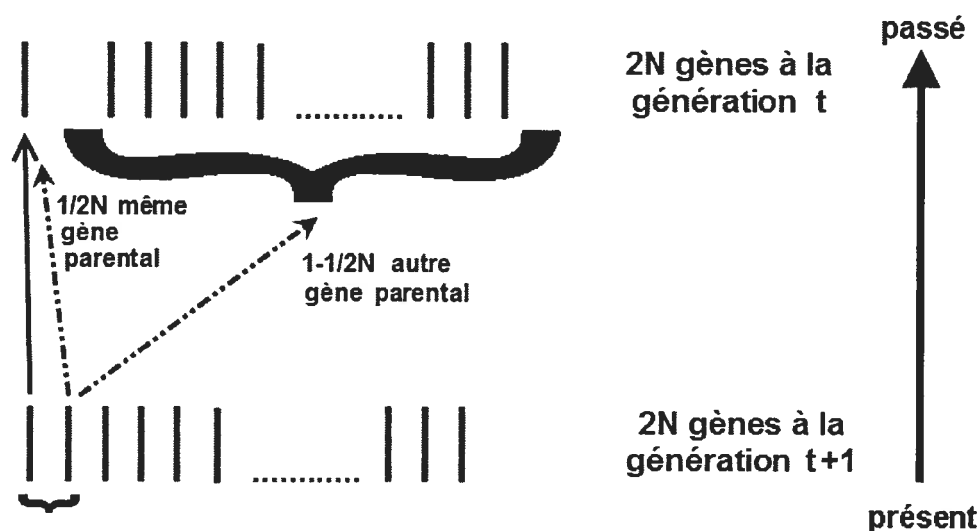


FIG. 1.1. Probabilités pour les gènes parentaux de deux gènes de la génération $t + 1$.

Soit T_2 le temps jusqu'à la coalescence de deux gènes en nombre de $2N$ générations (donc si $T_2 = 3$ alors T_2 correspond à $6N$ générations). Cette période de temps est intéressante, car c'est celle-ci qui sépare les deux gènes de leur ancêtre commun le plus récent. On peut donc associer cette période à une mesure de divergence entre les deux gènes. Pour tout $t > 0$, on a

$$\begin{aligned}
Pr(T_2 > t) \\
&= Pr(\text{nombre de g n rations jusqu'  coalescence   partir de 2 g nes} > [2Nt]) , \\
&= \left(1 - \frac{1}{2N}\right)^{[2Nt]} \xrightarrow{N \rightarrow \infty} e^{-t} ,
\end{aligned}$$

o  $[2Nt]$ d signe la valeur enti re de $2Nt$. Donc, T_2 suit asymptotiquement une loi exponentielle de param tre 1.

Maintenant, soit T_k le temps jusqu'  la premi re coalescence dans un ensemble de k g nes en nombre de $2N$ g n rations. Pour trouver la distribution de T_k , on doit passer par plusieurs  tapes. Trouvons premi rement la probabilit  que k g nes aient $k - 1$ parents. On a

$$\begin{aligned}
Pr(k \text{ g nes aient } k - 1 \text{ parents}) \\
&= Pr(\text{seulement deux des } k \text{ g nes coalescent en une g n ration}) , \\
&= \binom{k}{2} \frac{1}{2N} \left(1 - \frac{1}{2N}\right) \left(1 - \frac{2}{2N}\right) \dots \left(1 - \frac{k-2}{2N}\right) , \\
&= \frac{\binom{k}{2}}{2N} + o(1/N) , \tag{1.2.1}
\end{aligned}$$

o  $o(1/N)$ est une fonction petit ordre de $1/N$. Maintenant, la probabilit  que k g nes aient j parents, pour $1 \leq j \leq k - 2$, est

$$Pr(k \text{ g nes aient } j \text{ parents}) = 1 \left(1 - \frac{1}{2N}\right) \left(1 - \frac{2}{2N}\right) \dots \left(1 - \frac{j-1}{2N}\right) \frac{1}{(2N)^{k-j}} S_k^{(j)} ,$$

o  $S_k^{(j)}$ est le nombre de fa ons de r partir k g nes distincts en j groupes non-vides. C'est un nombre de Stirling de deuxi me esp ce. En d finissant r_i comme le nombre de g nes de la g n ration suivante provenant du parent i , on a

$$S_k^{(j)} = \frac{1}{j!} \sum_{\substack{r_1, r_2, \dots, r_j \geq 1 \\ r_1 + r_2 + \dots + r_j = k}} \binom{k}{r_1, r_2, \dots, r_j}.$$

On peut transformer la somme ci-dessus ainsi

$$\sum_{\substack{r_1, r_2, \dots, r_j \geq 1 \\ r_1 + r_2 + \dots + r_j = k}} \binom{k}{r_1, r_2, \dots, r_j} = \lambda_0 - \lambda_1 + \lambda_2 - \dots + (-1)^{j-1} \lambda_{j-1}, \quad (1.2.2)$$

où

$$\begin{aligned} \lambda_0 &= j^k, \quad \lambda_1 = \binom{j}{1} (j-1)^k, \quad \lambda_2 = \binom{j}{2} (j-2)^k, \dots \\ \dots, \quad \lambda_{j-1} &= \binom{j}{j-1} (j - (j-1))^k = j. \end{aligned}$$

Le membre de droite de l'équation (1.2.2) est obtenu en considérant qu'il peut y avoir des groupes r_i vides. En effet, le terme $\lambda_0 = j^k$ du membre de droite est le nombre de façons de répartir k gènes distincts en j groupes (vides ou non) alors que λ_1 est le nombre de façons de répartir k gènes distincts en $(j-1)$ groupes (vides ou non) lorsqu'il y a un groupe vide (un des j groupes). De façon semblable, λ_2 est le nombre de façons de répartir k gènes distincts en $(j-2)$ groupes (vides ou non) lorsqu'il y a deux groupes vides (deux des j groupes) et ainsi de suite jusqu'au terme λ_{j-1} qui est le nombre de façons de répartir k gènes distincts en 1 groupe (nécessairement non-vide) lorsqu'il y a $(j-1)$ groupes vides ($j-1$ des j groupes). Donc, pour obtenir le nombre de façons de répartir k gènes distincts en j groupes non-vides (c'est le membre de gauche de l'équation (1.2.2)), on doit donc soustraire λ_1 de λ_0 . Cependant, on soustrait alors deux fois λ_2 . On doit donc ajouter une fois λ_2 mais ce faisant, on ajoute une fois de trop λ_3 . Il faut donc soustraire λ_3 et on continue ainsi de suite jusqu'à λ_{j-1} . En répétant cette procédure d'inclusion-exclusion, on obtient

$$S_k^{(j)} = \frac{1}{j!} \left(j^k - \binom{j}{1} (j-1)^k + \binom{j}{2} (j-2)^k - \dots + (-1)^{j-1} j \right),$$

$$\leq \frac{j^k}{j!}.$$

Par conséquent, on a

$$Pr(k \text{ gènes aient } j \text{ parents}) \leq 1 \left(1 - \frac{1}{2N}\right) \left(1 - \frac{2}{2N}\right) \dots \left(1 - \frac{j-1}{2N}\right) \frac{1}{(2N)^{k-j}} \frac{j^k}{j!},$$

$$= o(1/N),$$
(1.2.3)

pour $1 \leq j \leq k-2$.

En utilisant les probabilités (1.2.1) et (1.2.3), on obtient la probabilité suivante :

$$Pr(k \text{ gènes aient } k \text{ parents}) = 1 - \frac{\binom{k}{2}}{2N} + o(1/N).$$

Avec cette dernière probabilité, il est possible de trouver la distribution limite de T_k lorsque $N \rightarrow \infty$, car

$$Pr(T_k > t) = Pr(\text{nbre gén. jusqu'à 1-ère coalescence à partir de } k \text{ gènes} > [2Nt]),$$

$$= \left(1 - \frac{\binom{k}{2}}{2N} + o(1/N)\right)^{[2Nt]} \xrightarrow{N \rightarrow \infty} e^{-\binom{k}{2}t}.$$

Donc on a que T_k suit asymptotiquement une loi exponentielle de paramètre $\binom{k}{2} = \frac{k(k-1)}{2}$, représentée par $Exp\left(\frac{k(k-1)}{2}\right)$. Maintenant qu'on a les distributions asymptotiques des T_k , on peut regarder la forme d'un arbre de coalescence pour disons cinq gènes (voir la figure 1.2). On débute initialement avec cinq gènes et on suit les lignées ancestrales jusqu'à la première coalescence (T_5 sera le temps

jusqu'à la première coalescence) pour se retrouver avec quatre gènes. On continue jusqu'à la deuxième coalescence (T_4 sera le temps entre la première et la deuxième coalescence) et on se retrouve avec trois gènes. On continue ainsi à remonter le temps jusqu'à l'ancêtre commun le plus récent (MRCA pour "most recent common ancestor") des cinq gènes. L'arbre de coalescence est donc associé à une mesure de divergence entre les gènes de l'échantillon.

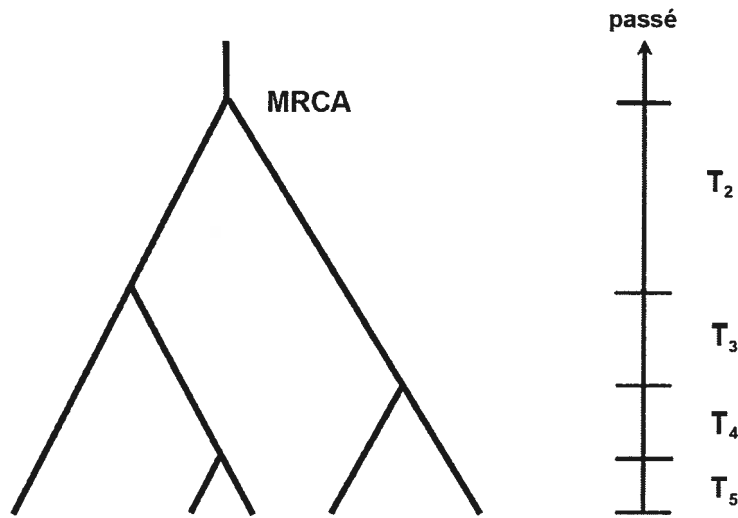


FIG. 1.2. Arbre de coalescence pour cinq gènes.

En connaissant la distribution des temps entre chaque coalescence, on peut facilement calculer le temps jusqu'au MRCA. On définit W_n le temps en nombre de $2N$ générations jusqu'au MRCA à partir de n gènes et on fait tendre n vers l'infini. Alors on a que $W_n = \sum_{k=2}^n T_k$ et que

$$\begin{aligned} E(W_n) &= \sum_{k=2}^n E(T_k) = \sum_{k=2}^n \frac{2}{k(k-1)} = 2 \sum_{k=2}^n \left(\frac{1}{k-1} - \frac{1}{k} \right), \\ &= 2 \left(1 - \frac{1}{n} \right) \xrightarrow{n \rightarrow \infty} 2. \end{aligned}$$

Donc lorsque n tend vers l'infini, le temps moyen avant d'atteindre le MRCA est 2. Il est intéressant de le comparer au temps moyen jusqu'au MRCA à partir de deux gènes (W_2). Puisque $W_2 = T_2$, alors

$$E(W_2) = E(T_2) = E(\text{Exp}(1)) = 1 .$$

C'est la moitié de l'espérance du temps de W_n lorsque n tend vers l'infini!! En effet, plus il y a de gènes, plus rapidement il y aura une coalescence et la décroissance, $E(T_2) > E(T_3) > \dots > E(T_{n-1}) > E(T_n)$, est très rapide.

Ce qui a été considéré jusqu'à présent n'est pas le processus de coalescence mais plutôt un processus de mort associé à ce dernier. Dans le processus de coalescence, on ne s'intéresse pas seulement au nombre de gènes ancestraux à différents moments mais aussi à la descendance des individus, c'est-à-dire qu'on doit savoir de qui descend chaque individu ($[T]$). Pour un processus avec n individus et à chaque génération t , on fait cela en définissant une relation d'équivalence entre les n individus. Deux individus sont dans la même classe d'équivalence s'ils ont un ancêtre commun à la génération t . Le nombre de classes d'équivalence représente donc le nombre d'ancêtres à la génération t et les individus d'une même classe sont tous des descendants de l'ancêtre représenté par la classe. À la génération dans laquelle l'échantillon de n individus a été pris, il y a donc n classes d'un seul individu ($\{\{1\}, \{2\}, \dots, \{n\}\}$) alors qu'à la génération du MRCA, il n'y a plus qu'une seule classe qui contient tous les individus ($\{\{1, 2, \dots, n\}\}$). Le processus de coalescence est donc un processus sur les partitions d'un ensemble.

La théorie de la coalescence peut être étendue à des populations avec sélection, mutation et recombinaison (voir [BBC]). Dans notre cas, elle nous sera plus tard

d'une grande utilité pour trouver les probabilités des différents génotypes pour un individu.

Chapitre 2

SÉLECTION ET PROBABILITÉ DE FIXATION

2.1. MODÈLE DE SÉLECTION

On a vu le modèle neutre de Wright-Fisher au chapitre 1.1 dans lequel on ne considère pas l'avantage sélectif que peut présenter un type de gène par rapport à un autre. Dans beaucoup de cas, les modèles avec sélection conviennent mieux à la réalité. En ajoutant la sélection au modèle du chapitre 1.1, c'est-à-dire si la valeur sélective (viabilité) des individus est différente pour chaque génotype, le tirage des N paires de gamètes est modifié. Chaque individu produit encore un nombre infini de gamètes qui s'unissent au hasard mais les proportions des génotypes sont par la suite modifiées proportionnellement aux valeurs sélectives qui affectent la viabilité. Si q_{ij} est la fréquence du génotype ij (ij peut être AA , AB ou BB) avant la sélection et si w_{ij} est la valeur sélective du génotype ij , alors la proportion de celui-ci après la sélection est

$$\tilde{q}_{ij} = \frac{q_{ij} \times w_{ij}}{(q_{AA} \times w_{AA}) + (q_{AB} \times w_{AB}) + (q_{BB} \times w_{BB})} \quad .$$

On se retrouve donc avec

$$\tilde{x} = \frac{(q_{AA} \times w_{AA}) + (1/2)(q_{AB} \times w_{AB})}{(q_{AA} \times w_{AA}) + (q_{AB} \times w_{AB}) + (q_{BB} \times w_{BB})}$$

comme nouvelle fréquence de A dans la population après la sélection. Le nombre de gènes de type A à la génération suivante suit alors une loi binomiale $(2N, \tilde{x})$ alors que les nombres d'individus de génotype AA, AB, BB , respectivement, suivent une loi multinomiale de paramètre $(N, \tilde{q}_{AA}, \tilde{q}_{AB}, \tilde{q}_{BB})$ telle que la probabilité qu'il y ait l individus de type AA , m de type AB , et n de type BB est

$$\binom{N}{l, m, n} (\tilde{q}_{AA})^l (\tilde{q}_{AB})^m (\tilde{q}_{BB})^n ,$$

où $l, m, n \geq 0$ et $l + m + n = N$.

2.2. APPROXIMATION DE LA PROBABILITÉ DE FIXATION PAR UN PROCESSUS DE DIFFUSION

On considère le modèle de Wright-Fisher appliqué à une population de N individus diploïdes (donc $2N$ gènes). Les générations sont discrètes et ne se chevauchent pas et à chaque génération, il y a N individus. On observe un locus qui a seulement deux allèles, A et B , et on suppose qu'il n'y a pas de mutation. La fréquence de l'allèle A peut prendre les valeurs $0/2N, 1/2N, 2/2N, \dots, (2N - 1)/2N, 2N/2N$ à chaque génération et si une des valeurs extrêmes ($0/2N$ ou $2N/2N$) est atteinte, la fréquence de A ne change plus par la suite. En effet, si cette fréquence atteint $0/2N$, il y a extinction de l'allèle A et il ne pourra pas revenir dans la population. Si la fréquence atteint $2N/2N$, il y a extinction de l'allèle B qui ne pourra pas revenir dans la population et donc fixation de l'allèle A . Ayant une population de taille fixe et finie, la fréquence de A sera égale à l'une ou l'autre de ces deux valeurs lorsque le nombre de générations tendra vers l'infini (voir la démonstration à la page 25). La probabilité de fixation de l'allèle A est la probabilité que la fréquence de A soit égale à $2N/2N$ lorsque le nombre de générations tend vers l'infini. C'est donc la probabilité qu'elle atteigne la valeur $2N/2N$ avant la valeur $0/2N$.

Pour trouver la probabilité de fixation, on utilisera une méthode directe en passant par la théorie de la coalescence (voir les sections 2.3 et 3.2.1). Cependant, avant de se lancer dans cette méthode directe, il est bien d'examiner un peu une autre méthode qui est plus souvent utilisée (voir [Le(2)]) : l'approximation par un processus de diffusion. Voici la définition du processus de diffusion ([KT]) :

Définition 2.2.1. *Un processus de diffusion X_t est un processus de Markov à temps continu et à trajectoires continues presque sûrement.*

Pour un petit Δt positif et étant donné $X_t = x$, le changement $\Delta X_t = X_{t+\Delta t} - X_t$ doit avoir une distribution de moyenne

$$E(\Delta X_t | X_t = x) = z(x)\Delta t + o(\Delta t) , \quad (2.2.1)$$

et il faut aussi que

$$E[(\Delta X_t)^2 | X_t = x] = v(x)\Delta t + o(\Delta t) . \quad (2.2.2)$$

Souvent, le changement ΔX_t satisfait aussi la condition suivante :

$$E[|(\Delta X_t)|^r | X_t = x] = o(\Delta t), \quad r = 3, 4, 5, \dots \quad (2.2.3)$$

On définit parfois un processus de diffusion comme un processus de Markov tel que les conditions (2.2.1) et (2.2.2) sont satisfaites et tel que l'équation (2.2.3) est valide pour au moins un $r > 2$. La fonction $z(x)$ est appelée la moyenne infinitésimale ou fonction de dérive alors que $v(x)$ est appelée la variance infinitésimale ou fonction de diffusion.

Dans le cas qui nous intéresse, le processus X_t représente l'approximation par diffusion de la fréquence de l'allèle A dans la population. Pour y arriver, on passe d'un processus à temps discret à un autre à temps continu en prenant $2N$ générations comme unité de temps (donc $\Delta t = 1/2N$) et en faisant tendre N vers l'infini (donc $\Delta t = 1/2N$ tend vers 0 et devient δt). Maintenant qu'on a un processus

de diffusion $X(t)$, il faut trouver comment exprimer la probabilité de fixation de l'allèle A pour un cas homogène (le processus ne dépend pas de t , il dépend de Δt) comme le nôtre.

Soit $F_A(x, t)$ la probabilité que l'allèle A se fixe dans la population après un temps t sachant que sa fréquence initiale $X(0)$ est x . Comme $X(t)$ est markovien, on a (voir [K])

$$\begin{aligned} F_A(x, t + \delta t) &= \int f(x, x + \delta x, \delta t) F_A(x + \delta x, t) d(\delta x) , \\ &\approx \int f(x, x + \delta x, \delta t) \left[F_A(x, t) + \frac{\partial F_A(x, t)}{\partial x} (\delta x) \right. \\ &\quad \left. + \frac{\partial^2 F_A(x, t)}{\partial x^2} \frac{(\delta x)^2}{2!} + \frac{\partial^3 F_A(x, t)}{\partial x^3} \frac{(\delta x)^3}{3!} \right] d(\delta x) , \end{aligned}$$

où $f(x, x + \delta x, \delta t)$ est la fonction de densité du passage de x à $x + \delta x$ pendant δt et γ est compris entre x et $x + \delta x$. En supposant qu'on peut sortir les dérivées partielles de l'intégrale, on a alors

$$\begin{aligned} F_A(x, t + \delta t) &= F_A(x, t) \int f(x, x + \delta x, \delta t) d(\delta x) \\ &\quad + \frac{\partial F_A(x, t)}{\partial x} \int f(x, x + \delta x, \delta t) (\delta x) d(\delta x) \\ &\quad + \frac{\partial^2 F_A(x, t)}{\partial x^2} \int f(x, x + \delta x, \delta t) \frac{(\delta x)^2}{2!} d(\delta x) \\ &\quad + \frac{\partial^3 F_A(x, t)}{\partial x^3} \int f(x, x + \delta x, \delta t) \frac{(\delta x)^3}{3!} d(\delta x) , \end{aligned}$$

$$\begin{aligned}
&= F_A(x, t) + \frac{\partial F_A(x, t)}{\partial x} \int f(x, x + \delta x, \delta t)(\delta x) d(\delta x) \\
&\quad + \frac{\partial^2 F_A(x, t)}{\partial x^2} \int f(x, x + \delta x, \delta t) \frac{(\delta x)^2}{2!} d(\delta x) \\
&\quad + \frac{\partial^3 F_A(x, t)}{\partial x^3} \int f(x, x + \delta x, \delta t) \frac{(\delta x)^3}{3!} d(\delta x) .
\end{aligned}$$

En mettant $F_A(x, t)$ à gauche de l'équation, en divisant par δt et en prenant la limite lorsque $\delta t \rightarrow 0$, on trouve

$$\begin{aligned}
\frac{\partial F_A(x, t)}{\partial t} &= \lim_{\delta t \rightarrow 0} \frac{F_A(x, t + \delta t) - F_A(x, t)}{\delta t} , \\
&= \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \left(\frac{\partial F_A(x, t)}{\partial x} \int f(x, x + \delta x, \delta t)(\delta x) d(\delta x) \right. \\
&\quad + \frac{\partial^2 F_A(x, t)}{\partial x^2} \int f(x, x + \delta x, \delta t) \frac{(\delta x)^2}{2!} d(\delta x) \\
&\quad \left. + \frac{\partial^3 F_A(x, t)}{\partial x^3} \int f(x, x + \delta x, \delta t) \frac{(\delta x)^3}{3!} d(\delta x) \right) , \\
&= z(x) \frac{\partial F_A(x, t)}{\partial x} + \frac{v(x)}{2} \frac{\partial^2 F_A(x, t)}{\partial x^2} , \tag{2.2.4}
\end{aligned}$$

en négligeant le terme dans lequel δx est d'ordre supérieur à deux, où

$$z(x) = \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} E(\delta x|x) = \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \int f(x, x + \delta x, \delta t)(\delta x) d(\delta x) ,$$

et

$$v(x) = \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} E[(\delta x)^2|x] = \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \int f(x, x + \delta x, \delta t)(\delta x)^2 d(\delta x) .$$

Comme on veut trouver la probabilité de fixation éventuelle, c'est-à-dire $F_A(x) = \lim_{t \rightarrow \infty} F_A(x, t)$ (donc $\frac{\partial F_A(x, t)}{\partial t} = 0$), l'équation (2.2.4) à résoudre devient l'équation différentielle ordinaire

$$z(x) \frac{dF_A(x)}{dx} + \frac{v(x)}{2} \frac{d^2 F_A(x)}{dx^2} = 0 , \quad (2.2.5)$$

avec $F_A(0) = 0$ et $F_A(1) = 1$ comme conditions aux frontières. En posant $\psi(x) = \frac{dF_A(x)}{dx}$ et en réorganisant les termes de (2.2.5) comme suit

$$\frac{\psi'(x)}{\psi(x)} = -2 \frac{z(x)}{v(x)} ,$$

ce qui implique

$$\int \frac{d}{dx} \log[\psi(x)] dx = -2 \int \frac{z(x)}{v(x)} dx ,$$

on déduit que

$$F_A(x) = \frac{\int_0^x \psi(y) dy}{\int_0^1 \psi(y) dy} , \quad (2.2.6)$$

où

$$\psi(y) = \exp \left(-2 \int^y \frac{z(x)}{v(x)} dx \right) . \quad (2.2.7)$$

La forme de la probabilité de fixation (équation (2.2.6)) permet de laisser indéfinie l'intégrale de l'équation (2.2.7). En effet avec une intégrale définie, il y a

apparition d'une constante qui s'annule dû à la division dans (2.2.6).

Donnons maintenant un exemple d'approximation par le processus de diffusion. Soit la population diploïde décrite au début de cette section. Supposons que les génotypes n'ont pas tous le même avantage sélectif, plus précisément que les valeurs sélectives sont les suivantes :

$$w_{AA} = 1 + s u_{AA} ,$$

$$w_{AB} = 1 + s u_{AB} ,$$

$$w_{BB} = 1 + s u_{BB} ,$$

où s est d'ordre plus petit que $1/2N$ (disons $s = d/2N$ avec d petit) et représente l'intensité de la sélection. Soit $1/2N$ la fréquence initiale de A dont on veut la probabilité de fixation éventuelle. On compte le temps en nombre de $2N$ générations et alors une génération équivaut à $\Delta t = 1/2N$. Par la suite, on obtient $z(x)$ et $v(x)$ par les équations (2.2.1) et (2.2.2). Il nous faut donc calculer l'espérance et le deuxième moment de $\Delta X_t = X_{t+\Delta t} - X_t$ étant donné que $X_t = x$. Pour l'espérance, on trouve

$$\begin{aligned} E(X_{t+\Delta t} - X_t | X_t = x) &= E(X_{t+\Delta t} | X_t = x) - x , \\ &= \frac{x^2 w_{AA} + (1/2) 2x(1-x) w_{AB}}{x^2 w_{AA} + 2x(1-x) w_{AB} + (1-x)^2 w_{BB}} - x , \\ &= \frac{x^2 (1 + s u_{AA}) + (1/2) 2x(1-x) (1 + s u_{AB})}{1 + x^2 (s u_{AA}) + 2x(1-x) (s u_{AB}) + (1-x)^2 (s u_{BB})} - x . \end{aligned}$$

Le premier terme à droite de la dernière égalité est de forme $a/(1-r)$, où $|r| < 1$ pour s suffisamment petit. On peut donc le transformer en série géométrique : $a \sum_{i=0}^{\infty} r^i$. On obtient alors

$$\begin{aligned}
E(\Delta X_t | X_t = x) &= [x^2 (1 + su_{AA}) + (1/2) 2x(1-x) (1 + su_{AB})] \\
&\quad \times [1 - x^2(su_{AA}) - 2x(1-x)(su_{AB}) - (1-x)^2(su_{BB}) + o(s)] - x , \\
&= [x^2(1-x)u_{AA} + x(1-x)(1-2x)u_{AB} - x(1-x)^2u_{BB}]s + o(s) .
\end{aligned}
\tag{2.2.8}$$

Puisque $s = d/2N$ et $\Delta t = 1/2N$, on conclut, en comparant les équations (2.2.1) et (2.2.8), que

$$z(x) = dx(1-x)[xu_{AA} + (1-2x)u_{AB} - (1-x)u_{BB}] .$$

Pour trouver $v(x)$, on calcule la variance de ΔX_t étant donné que $X_t = x$, puisque

$$\begin{aligned}
Var(\Delta X_t | X_t = x) &= E((\Delta X_t)^2 | X_t = x) - E^2(\Delta X_t | X_t = x) , \\
&= E((\Delta X_t)^2 | X_t = x) - o(s) ,
\end{aligned}$$

et alors

$$Var(\Delta X_t | X_t = x) = v(x)/2N + o(s).$$

On trouve

$$\begin{aligned}
Var(\Delta X_t | X_t = x) &= Var(X_{t+\Delta t} - X_t | X_t = x) , \\
&= Var(X_{t+\Delta t} | X_t = x) + Var(-X_t | X_t = x) \\
&\quad + 2Cov(X_{t+\Delta t}, -X_t | X_t = x) , \\
&= Var(X_{t+\Delta t} | X_t = x) , \\
&= Var\left(Q_{t+\Delta t}^{AA} + \frac{1}{2}Q_{t+\Delta t}^{AB} \middle| X_t = x\right) ,
\end{aligned}$$

où Q_t^{AA} par exemple désigne la fréquence du génotype AA à la génération t . On a alors

$$\begin{aligned}
&Var\left(Q_{t+\Delta t}^{AA} + \frac{1}{2}Q_{t+\Delta t}^{AB} \middle| X_t = x\right) \\
&= Var\left(Q_{t+\Delta t}^{AA} \middle| X_t = x\right) + Var\left(\frac{1}{2}Q_{t+\Delta t}^{AB} \middle| X_t = x\right) + 2Cov\left(Q_{t+\Delta t}^{AA}, \frac{1}{2}Q_{t+\Delta t}^{AB} \middle| X_t = x\right) , \\
&= \frac{1}{N^2} Var\left(n_{t+\Delta t}^{AA} \middle| X_t = x\right) + \frac{1}{4N^2} Var\left(n_{t+\Delta t}^{AB} \middle| X_t = x\right) \\
&\quad + \frac{1}{N^2} Cov\left(n_{t+\Delta t}^{AA}, n_{t+\Delta t}^{AB} \middle| X_t = x\right) ,
\end{aligned}$$

où $n_{t+\Delta t}^{ij}$ représente le nombre d'individus de génotype ij à la génération $t + \Delta t$. On a vu à la section 2.1 que les nombres d'individus de génotype AA, AB, BB , respectivement, suivent une loi multinomiale de paramètre $(N, \tilde{q}_t^{AA}, \tilde{q}_t^{AB}, \tilde{q}_t^{BB})$, où \tilde{q}_t^{ij} est la fréquence des individus de génotype ij après la sélection juste avant le tirage des N individus pour former la génération $t + \Delta t$. On en conclut que

$$Var\left(Q_{t+\Delta t}^{AA} + \frac{1}{2}Q_{t+\Delta t}^{AB} \middle| X_t = x\right) = \frac{N\tilde{q}_t^{AA}(1 - \tilde{q}_t^{AA})}{N^2} + \frac{N\tilde{q}_t^{AB}(1 - \tilde{q}_t^{AB})}{4N^2} - \frac{\tilde{q}_t^{AA}\tilde{q}_t^{AB}N}{N^2} .$$

Puisque

$$\begin{aligned}\tilde{q}_t^{AA} &= \frac{x^2 w_{AA}}{x^2 w_{AA} + 2x(1-x)w_{AB} + (1-x)^2 w_{BB}} , \\ &= x^2 + O(s) ,\end{aligned}$$

et

$$\begin{aligned}\tilde{q}_t^{AB} &= \frac{2x(1-x) w_{AB}}{x^2 w_{AA} + 2x(1-x)w_{AB} + (1-x)^2 w_{BB}} , \\ &= 2x(1-x) + O(s) ,\end{aligned}$$

avec $s = d/2N$, on a finalement

$$\begin{aligned}Var(\Delta X_t | X_t = x) &= \frac{x^2(1-x^2) + (1/2)x(1-x) - x^2(1-x)^2 - 2x^3(1-x)}{N} + o(1/N) , \\ &= \frac{x(1-x)}{2N} + o(1/N) ,\end{aligned}\tag{2.2.9}$$

d'où

$$v(x) = x(1-x) .$$

La prochaine étape est d'évaluer la fonction $\psi(y)$ (voir (2.2.7)). On trouve

$$\begin{aligned}\psi(y) &= \exp \left(- 2 \int^y \frac{z(x)}{v(x)} dx \right) , \\ &= \exp \left(- 2d \int^y [xu_{AA} + (1-2x)u_{AB} - (1-x)u_{BB}] dx \right) , \\ &= \exp \left(- 2d \left(\frac{y^2}{2} u_{AA} + y(1-y)u_{AB} - \frac{y}{2}(2-y)u_{BB} \right) \right) .\end{aligned}$$

Il ne reste maintenant qu'à trouver la probabilité de fixation (voir (2.2.6)). Cependant, pour arriver facilement à cette probabilité, on développe $\psi(y)$ en série de Taylor pour ensuite intégrer les termes d'ordre 1 ou d et négliger les autres. C'est une bonne approximation, car on suppose que d est petit. On a donc

$$\begin{aligned}
 F_A(x) &= \frac{\int_0^x \psi(y) dy}{\int_0^1 \psi(y) dy} , \\
 &\approx \frac{\int_0^x \left[1 - 2d \left(\frac{y^2}{2} u_{AA} + y(1-y) u_{AB} - \frac{y}{2} (2-y) u_{BB} \right) \right] dy}{\int_0^1 \left[1 - 2d \left(\frac{y^2}{2} u_{AA} + y(1-y) u_{AB} - \frac{y}{2} (2-y) u_{BB} \right) \right] dy} , \\
 &= \frac{x - 2d \left(\frac{x^3}{6} u_{AA} + x^2 \left(\frac{1}{2} - \frac{x}{3} \right) u_{AB} - x^2 \left(\frac{1}{2} - \frac{x}{6} \right) u_{BB} \right)}{1 - 2d \left(\frac{1}{6} u_{AA} + \frac{1}{6} u_{AB} - \frac{1}{3} u_{BB} \right)} .
 \end{aligned}$$

On a encore une expression de la forme $a/(1-r)$ avec $|r| < 1$. On l'écrit sous la forme $a \sum_{i=0}^{\infty} r^i$, et on trouve finalement

$$\begin{aligned}
 F_A(x) &= \left[x - 2d \left(\frac{x^3}{6} u_{AA} + x^2 \left(\frac{1}{2} - \frac{x}{3} \right) u_{AB} - x^2 \left(\frac{1}{2} - \frac{x}{6} \right) u_{BB} \right) \right] \\
 &\quad \times \left[1 + 2d \left(\frac{1}{6} u_{AA} + \frac{1}{6} u_{AB} - \frac{1}{3} u_{BB} \right) + o(d) \right] , \\
 &= x + 2dx \left[\left(\frac{1}{6} - \frac{x^2}{6} \right) u_{AA} + \left(\frac{1}{6} - \frac{x}{2} + \frac{x^2}{3} \right) u_{AB} + \left(-\frac{1}{3} + \frac{x}{2} - \frac{x^2}{6} \right) u_{BB} \right] + o(d) .
 \end{aligned}
 \tag{2.2.10}$$

L'approximation par le processus de diffusion fonctionne bien si on considère une population de taille très grande. Avec une population de petite taille, il faut plutôt utiliser la méthode de la section suivante, plus directe.

2.3. APPROCHE DIRECTE POUR LA PROBABILITÉ DE FIXATION PAR LA THÉORIE DE LA COALESCENCE

La probabilité de fixation d'un allèle A dépend de sa fréquence à la génération initiale et de sa valeur sélective. On suppose qu'il y a seulement un gène de type A dans la population à la génération initiale ($X(0) = x_0 = 1/2N$). La probabilité de fixation de A peut alors s'exprimer en fonction de l'intensité de la sélection s uniquement ($F_A(s)$). Si elle est continue et deux fois dérivable en s , on a

$$F_A(s) = F_A(0) + sF'_A(0) + o(s) = x_0 + sF'_A(0) + o(s), \quad (2.3.1)$$

où $o(s)$ désigne une fonction petit ordre de s . Montrons d'abord que dans une population diploïde de taille $2N$ suivant le modèle de Wright-Fisher sans mutation et sans sélection, la probabilité de fixation de A est sa fréquence initiale, c'est-à-dire, $F_A(0) = x_0$.

Démonstration. Soit Y_n le nombre de gènes de type A à la génération n . On a donc $Y_0 = x_0 \times 2N = 1$. On suppose qu'il n'y a pas de sélection. Dans ce cas, à chacun des $2N$ tirages pour former la nouvelle génération, la probabilité de piger un gène de type A est exactement égale à la fréquence de A à la génération courante. Pour un m particulier, si $Y_m = y_m$, alors Y_{m+1} suit une loi binomiale $(2N, y_m/2N)$. Il s'ensuit que $\{Y_n\}_{n \geq 0}$ est une martingale bornée, car $0 \leq Y_n \leq 2N$ pour $n \geq 0$ et

$$E(Y_{m+1} | Y_m = y_m) = 2N \times \frac{y_m}{2N} = y_m.$$

Soit maintenant T le temps d'arrêt défini par $T = \min\{n : Y_n = 0 \text{ ou } 2N\}$. On suppose pour l'instant que $Pr(T < \infty) = 1$ et on le prouvera par la suite.

On peut alors utiliser le théorème d'arrêt pour les martingales bornées (voir [Du]) qui garantit que $E(Y_T) = E(Y_0)$. En explicitant, on trouve que

$$x_0 \times 2N = E(Y_0) = E(Y_T) = 2N \times F_A(0) + 0 \times (1 - F_A(0)) = 2N \times F_A(0) .$$

Il s'ensuit que $F_A(0) = x_0$.

□

Il reste à montrer que la fréquence de A sera égale soit à 0, soit à 1 lorsque le nombre de générations tendra vers l'infini, c'est-à-dire que $Pr(T < \infty) = 1$.

Démonstration. Soit la même situation que dans la démonstration précédente avec les mêmes définitions pour Y_n et T . On rappelle qu'il n'y a pas de sélection. On sait que $T > 0$, car $Y_0 = 1$, et on suppose $T > m$, où $m \geq 0$ est fixé et fini. Il y a trois cas à considérer.

Cas 1.

Si $Y_m = N$, alors

$$Pr(Y_{m+1} = 2N | Y_m = N) = \left(\frac{N}{2N}\right)^{2N} = \left(\frac{1}{2}\right)^{2N} ,$$

et donc

$$Pr(T \leq m+1 | Y_m = N, T > m) = Pr(T = m+1 | Y_m = N, T > m) \geq \left(\frac{1}{2}\right)^{2N} ,$$

car

$$Pr(T = m + 1 | Y_m = N, T > m) = Pr(Y_{m+1} = 2N \text{ ou } 0 | Y_m = N) .$$

Cas 2.

Maintenant, si $2N > Y_m > N$, alors

$$Pr(Y_{m+1} = 2N | Y_m > N) > \left(\frac{N}{2N}\right)^{2N} = \left(\frac{1}{2}\right)^{2N} ,$$

et donc

$$Pr(T \leq m + 1 | Y_m > N, T > m) = Pr(T = m + 1 | Y_m > N, T > m) > \left(\frac{1}{2}\right)^{2N} .$$

Cas 3.

Enfin, si $0 < Y_m < N$, alors

$$Pr(Y_{m+1} = 0 | Y_m < N) > \left(\frac{N}{2N}\right)^{2N} = \left(\frac{1}{2}\right)^{2N} ,$$

et donc

$$Pr(T \leq m + 1 | Y_m < N, T > m) = Pr(T = m + 1 | Y_m < N, T > m) > \left(\frac{1}{2}\right)^{2N} .$$

Dans tous les cas, on trouve que $Pr(T \leq m + 1 | T > m) \geq \left(\frac{1}{2}\right)^{2N} > 0$, et donc il y a une borne inférieure positive pour la probabilité qu'il y ait soit fixation, soit extinction de A à la prochaine génération (la génération $m + 1$) peu importe la valeur de $0 < Y_m < 2N$. On définit le temps d'arrêt $T_m = T - m$ sachant que $T > m$. On a donc que T_m prend des valeurs positives non-nulles et que $Pr(T_m \leq 1 | T > m) \geq \left(\frac{1}{2}\right)^{2N} > 0$. On peut maintenant utiliser le lemme du piéton (voir [Du]) et on a

$$Pr(T_m > n | T > m) \leq \left(1 - \left(\frac{1}{2}\right)^{2N}\right)^n .$$

En prenant la limite lorsque $n \rightarrow \infty$ et en utilisant le fait que $\left(1 - \left(\frac{1}{2}\right)^{2N}\right) < 1$, on trouve que

$$Pr(T_m = \infty) = 0 .$$

On a donc

$$\begin{aligned} Pr(T = \infty) &= Pr(T_m + m = \infty) , \\ &= 0 , \end{aligned}$$

car T_m et m sont positifs et finis, ce qui implique que

$$Pr(T < \infty) = 1 - Pr(T = \infty) = 1 .$$

□

On revient à l'équation (2.3.1). Il ne nous reste qu'à trouver $F'_A(0)$. Pour ce faire, nous allons utiliser une méthode directe et exacte tirée de [LL].

On considère une population de N individus diploïdes (donc $2N$ gènes) qui suit le modèle de Wright-Fisher avec sélection mais sans mutation. On suppose qu'il n'y a que deux allèles, A et B , au locus considéré, et que la fréquence initiale de A est $1/2N$. On suppose aussi que les valeurs sélectives sont de la forme

$$w_{AA} = 1 + s ,$$

$$w_{AB} = 1 + sh ,$$

$$w_{BB} = 1 ,$$

où $0 < h < 1$ et s représente l'intensité de la sélection. Pour trouver $F'_A(0)$, on exprime $F_A(s)$ d'une autre manière. On a vu plus tôt qu'à long terme, soit A se fixe (i.e., la fréquence de A est $2N/2N = 1$), soit il disparaît de la population (i.e., la fréquence de A est $0/2N = 0$). Si X_t est la fréquence de A à la génération t , on a donc que X_t converge en probabilité vers

$$X_\infty = \begin{cases} 1 & \text{avec probabilité } F_A(s), \\ 0 & \text{avec probabilité } 1 - F_A(s), \end{cases}$$

lorsque $t \rightarrow \infty$ (voir la démonstration un peu plus haut). Puisque $\{X_t\}_{t \geq 0}$ est uniformément borné, $\lim_{t \rightarrow \infty} E_s(X_t)$ existe et vaut $E_s(X_\infty) = F_A(s)$, où E_s est l'espérance dans le cas où l'intensité de la sélection est s . On peut écrire

$$E_s(X_\infty - X_0) = \lim_{k \rightarrow \infty} E_s(X_k - X_0) ,$$

ce qui implique que

$$\begin{aligned} E_s(X_\infty) - E_s(X_0) &= \lim_{k \rightarrow \infty} E_s \left[\sum_{t=0}^{k-1} (X_{t+1} - X_t) \right] , \\ &= E_s \left[\sum_{t=0}^{\infty} (X_{t+1} - X_t) \right] . \end{aligned}$$

Il est montré dans [LL] qu'on peut entrer l'espérance dans la somme et que la dérivée de la somme est la somme des dérivées pour un modèle plus général dans

des conditions de régularité qui sont ici vérifiées. Dans ce qui suit, on reprend les arguments présentés dans [LL] et on les applique au cas particulier qui nous occupe. En utilisant l'espérance conditionnelle, on a

$$\begin{aligned} E_s(X_\infty) - E_s(X_0) &= \sum_{t=0}^{\infty} E_s(X_{t+1} - X_t) , \\ &= \sum_{t=0}^{\infty} E_s\left(E_s(X_{t+1} - X_t | X_t)\right) , \\ &= \sum_{t=0}^{\infty} E_s(g_s(X_t)) , \end{aligned}$$

où $g_s(X_t) = E_s(X_{t+1} - X_t | X_t)$. Puisque $E_s(X_0) = x_0 = 1/2N$, on obtient que

$$F_A(s) = \frac{1}{2N} + \sum_{t=0}^{\infty} E_s(g_s(X_t)) . \quad (2.3.2)$$

On définit maintenant $D_{ij}(s)$ la probabilité que la fréquence de A passe de $i/2N$ à $j/2N$ en une génération. La matrice de transition D est telle que $D^t = D^{(t)}$, où $D^{(t)}$ est la matrice de transition dont les entrées sont les probabilités de passer d'un état à un autre en t générations. On a alors que

$$E_s(g_s(X_t)) = \sum_{j=0}^{2N} g_s\left(\frac{j}{2N}\right) D_{1j}^{(t)}(s) . \quad (2.3.3)$$

On déduit ensuite $F'_A(0)$ des équations (2.3.2) et (2.3.3). On dérive d'abord l'équation (2.3.2) et on obtient

$$\begin{aligned}
F'_A(s) &= \sum_{t=0}^{\infty} \frac{d}{ds} E_s(g_s(X_t)) , \\
&= \sum_{t=0}^{\infty} \frac{d}{ds} \left[\sum_{j=0}^{2N} g_s\left(\frac{j}{2N}\right) D_{1j}^{(t)}(s) \right] , \\
&= \sum_{t=0}^{\infty} \left[E'_s(g_s(X_t)) + E_s(g'_s(X_t)) \right] , \tag{2.3.4}
\end{aligned}$$

où

$$E'_s(g_s(X_t)) = \sum_{j=0}^{2N} g_s\left(\frac{j}{2N}\right) \frac{d}{ds} D_{1j}^{(t)}(s) , \tag{2.3.5}$$

et

$$E_s(g'_s(X_t)) = \sum_{j=0}^{2N} \frac{d}{ds} \left(g_s\left(\frac{j}{2N}\right) \right) D_{1j}^{(t)}(s) .$$

Puisque $g_s(X_t) = 0$ lorsque $s = 0$, on en tire que l'équation (2.3.5) est égale à 0 lorsque $s = 0$, et donc

$$F'_A(0) = \sum_{t=0}^{\infty} \sum_{j=0}^{2N} \left[\frac{d}{ds} \left(g_s\left(\frac{j}{2N}\right) \right) \right]_{s=0} D_{1j}^{(t)}(0) .$$

Puisqu'on a le même modèle qu'à la page 19, la valeur de $g_s(X_t)$ peut être trouvée facilement à partir de l'équation (2.2.8) en remplaçant u_{AA} , u_{AB} et u_{BB} respectivement par 1, h et 0. Si $X_t = x$, on a

$$\begin{aligned}
g_s(x) &= E_s(X_{t+1} - X_t | X_t = x) , \\
&= [x^2(1-x) + x(1-x)(1-2x)h]s + o(s) , \\
&= sx(1-x)(x+h-2hx) + o(s).
\end{aligned}$$

On obtient donc

$$\begin{aligned}
 F'_A(0) &= \sum_{t=0}^{\infty} \sum_{j=0}^{2N} \left[\frac{d}{ds} \left[\frac{j}{2N} \left(1 - \frac{j}{2N} \right) \left(\frac{j}{2N} + h - 2h \left(\frac{j}{2N} \right) \right) s + o(s) \right]_{s=0} \times D_{1j}^{(t)}(0) \right] , \\
 &= \sum_{t=0}^{\infty} \sum_{j=0}^{2N} \left[\frac{j}{2N} \left(1 - \frac{j}{2N} \right) \left(\frac{j}{2N} + h - 2h \left(\frac{j}{2N} \right) \right) D_{1j}^{(t)}(0) \right] , \\
 &= \sum_{t=0}^{\infty} E_0 \left(X_t (1 - X_t) (X_t + h - 2h X_t) \right) , \\
 &= \sum_{t=0}^{\infty} E_0 \left((X_t - X_t^2) (h + (1 - 2h) X_t) \right) , \\
 &= h \sum_{t=0}^{\infty} E_0 (X_t - X_t^2) + (1 - 2h) \sum_{t=0}^{\infty} E_0 (X_t^2 - X_t^3) . \tag{2.3.6}
 \end{aligned}$$

La prochaine étape est de trouver $\sum_{t=0}^{\infty} E_0 (X_t^i - X_t^{i+1})$ pour $i = 1, 2$. On peut exprimer X_t sous la forme

$$\begin{aligned}
 X_t &= \frac{1}{N} \sum_{l=1}^N \left(\frac{\xi_l + \eta_l}{2} \right) , \\
 &= \frac{1}{2N} \sum_{l=1}^N (\xi_l + \eta_l) ,
 \end{aligned}$$

où

$$\xi_l = \begin{cases} 1 & \text{si le gène provenant du père de l'individu } l \text{ à la génération } t \text{ est de type } A, \\ 0 & \text{sinon,} \end{cases}$$

et

$$\eta_l = \begin{cases} 1 & \text{si le gène provenant de la mère de l'individu } l \text{ à la génération } t \text{ est de type } A, \\ 0 & \text{sinon.} \end{cases}$$

On a alors que

$$\begin{aligned} X_t^i &= \frac{1}{(2N)^i} \sum_{\substack{i_1, i_2, \dots, i_N \geq 0 \\ i_1 + i_2 + \dots + i_N = i}} \left[\binom{i}{i_1, \dots, i_N} \prod_{l=1}^N (\xi_l + \eta_l)^{i_l} \right], \\ &= \frac{1}{(2N)^i} \sum_{\substack{i_1, i_2, \dots, i_N \geq 0 \\ i_1 + i_2 + \dots + i_N = i}} \left[\binom{i}{i_1, \dots, i_N} \prod_{l=1}^N \left(\sum_{k_l=0}^{i_l} \binom{i_l}{k_l} \xi_l^{k_l} \eta_l^{i_l - k_l} \right) \right], \\ &= \frac{1}{(2N)^i} \sum_{\substack{i_1, i_2, \dots, i_N \geq 0 \\ i_1 + i_2 + \dots + i_N = i}} \left[\binom{i}{i_1, \dots, i_N} \sum_{k_1=0}^{i_1} \sum_{k_2=0}^{i_2} \dots \sum_{k_N=0}^{i_N} \binom{i_1}{k_1} \dots \binom{i_N}{k_N} \xi_1^{k_1} \eta_1^{i_1 - k_1} \dots \xi_N^{k_N} \eta_N^{i_N - k_N} \right]. \end{aligned}$$

En prenant l'espérance de X_t^i lorsque $s = 0$, on obtient

$$\begin{aligned} E_0(X_t^i) &= \frac{1}{(2N)^i} \sum_{\substack{i_1, i_2, \dots, i_N \geq 0 \\ i_1 + i_2 + \dots + i_N = i}} \left[\binom{i}{i_1, \dots, i_N} \sum_{k_1=0}^{i_1} \dots \sum_{k_N=0}^{i_N} \binom{i_1}{k_1} \dots \binom{i_N}{k_N} E_0 \left(\xi_1^{k_1} \eta_1^{i_1 - k_1} \dots \xi_N^{k_N} \eta_N^{i_N - k_N} \right) \right], \\ &= \frac{1}{(2N)^i} \sum_{\substack{i_1, i_2, \dots, i_N \geq 0 \\ i_1 + i_2 + \dots + i_N = i}} \sum_{k_1=0}^{i_1} \dots \sum_{k_N=0}^{i_N} \left[\binom{i}{i_1, \dots, i_N} \binom{i_1}{k_1} \dots \binom{i_N}{k_N} E_0 \left(\xi_1^{k_1} \eta_1^{i_1 - k_1} \dots \xi_N^{k_N} \eta_N^{i_N - k_N} \right) \right], \\ &= \frac{1}{(2N)^i} \sum_{\substack{i_1, i_2, \dots, i_N \geq 0 \\ i_1 + i_2 + \dots + i_N = i}} \sum_{k_1=0}^{i_1} \dots \sum_{k_N=0}^{i_N} \left[\binom{i}{k_1, i_1 - k_1, \dots, k_N, i_N - k_N} \right. \\ &\quad \left. \times E_0 \left(\xi_1^{k_1} \eta_1^{i_1 - k_1} \dots \xi_N^{k_N} \eta_N^{i_N - k_N} \right) \right], \end{aligned}$$

et en posant $k_{N+1} = i_1 - k_1$, $k_{N+2} = i_2 - k_2$, ..., $k_{2N} = i_N - k_N$, on a

$$\begin{aligned}
E_0(X_t^i) &= \frac{1}{(2N)^i} \sum_{\substack{i_1, i_2, \dots, i_N \geq 0 \\ i_1 + i_2 + \dots + i_N = i}} \sum_{\substack{k_1, k_{N+1} \geq 0 \\ k_1 + k_{N+1} = i_1}} \dots \sum_{\substack{k_N, k_{2N} \geq 0 \\ k_N + k_{2N} = i_N}} \left[\binom{i}{k_1, \dots, k_{2N}} E_0(\xi_1^{k_1} \dots \xi_N^{k_N} \eta_1^{k_{N+1}} \dots \eta_N^{k_{2N}}) \right], \\
&= \frac{1}{(2N)^i} \sum_{\substack{k_1, k_2, \dots, k_{2N} \geq 0 \\ k_1 + k_2 + \dots + k_{2N} = i}} \left[\binom{i}{k_1, \dots, k_{2N}} E_0(\xi_1^{k_1} \dots \xi_N^{k_N} \eta_1^{k_{N+1}} \dots \eta_N^{k_{2N}}) \right].
\end{aligned}$$

La dernière équation peut se simplifier car les $2N$ variables ξ_n et η_n , $1 \leq n \leq N$ sont échangeables lorsque $s = 0$. En général, on dit que des variables Z_1, \dots, Z_{2N} sont échangeables si pour tout $1 \leq n \leq 2N$, la distribution de $(Z_{m_1}, \dots, Z_{m_n})$ est la même en dépit des permutations de variables présentes dans celle-ci ([Bi]). C'est le cas des $2N$ variables ξ_n et η_n , $1 \leq n \leq N$, dans le modèle neutre de Wright-Fisher dû au tirage au hasard. En notant

$$\varphi_n = \begin{cases} \xi_n & \text{si } n \leq N, \\ \eta_{n-N} & \text{si } n > N, \end{cases}$$

et puisque $\varphi_n^k = \varphi_n$ si $k \geq 1$ et 1 si $k = 0$, on trouve

$$E_0(X_t^i) = \frac{1}{(2N)^i} \sum_{l=1}^{\min(i, 2N)} \left[(2N)_{[l]} S_i^{(l)} E_0(\varphi_1 \dots \varphi_l) \right],$$

où $(2N)_{[l]} = 2N \times (2N-1) \times \dots \times (2N-l+1)$ et $S_i^{(l)}$ est un nombre de Stirling de deuxième espèce qui représente le nombre de façons de répartir i gènes distincts en l groupes non-vides, c'est-à-dire

$$S_i^{(l)} = \frac{1}{l!} \sum_{\substack{r_1, r_2, \dots, r_l \geq 1 \\ r_1 + r_2 + \dots + r_l = i}} \binom{i}{r_1, r_2, \dots, r_l}. \quad (2.3.7)$$

On a aussi

$$\begin{aligned}
 E_0(\varphi_1 \dots \varphi_l) &= 1 \times Pr_0(\varphi_1 = 1, \dots, \varphi_l = 1) + 0 \times (1 - Pr_0(\varphi_1 = 1, \dots, \varphi_l = 1)) , \\
 &= Pr_0(\varphi_1 = 1, \dots, \varphi_l = 1) , \\
 &= \frac{1}{2N} Pr_0(T_l \leq t) ,
 \end{aligned}$$

où T_l est le nombre de générations qu'il faut remonter avant d'arriver au MRCA à partir de l gènes à la génération t et Pr_0 désigne la probabilité lorsque $s = 0$. Le MRCA des l gènes, s'il existe, doit être de type A , car $\varphi_j = 1$ seulement si le gène j est de type A . La probabilité que le MRCA des l gènes soit de type A est $1/2N$, car c'est la fréquence de A initialement. On a alors

$$E_0(\varphi_1 \dots \varphi_l) = \frac{1}{2N} - \frac{1}{2N} Pr_0(T_l > t) ,$$

et donc

$$\begin{aligned}
 E_0(X_t^i) &= \frac{1}{(2N)^{i+1}} \sum_{l=1}^{\min(i, 2N)} \left[(2N)_{[l]} S_i^{(l)} \right] - \frac{1}{(2N)^{i+1}} \sum_{l=1}^{\min(i, 2N)} \left[(2N)_{[l]} S_i^{(l)} Pr_0(T_l > t) \right] .
 \end{aligned}$$

Le premier terme du membre de droite de l'équation ci-dessus peut être simplifié, car on a

$$\begin{aligned}
 \sum_{l=1}^{\min(i, 2N)} \left[(2N)_{[l]} S_i^{(l)} \right] &= \sum_{l=1}^{\min(i, 2N)} \left[\binom{2N}{l} (S_i^{(l)} \times l!) \right] , \\
 &= (2N)^i .
 \end{aligned}$$

En effet, $(2N)^i$ est le nombre de façons de placer i éléments dans $2N$ groupes (vides ou non) distincts. On peut aussi arriver à ce nombre en passant par des étapes bien spécifiques. On pose l étant le nombre de groupes non-vides. Pour chaque valeur de l , on doit alors compter le nombre de façons de choisir l groupes parmi $2N$ (c'est $\binom{2N}{l}$), et pour chacun de ces choix, on doit compter le nombre de façons de placer i éléments dans les l groupes non-vides distincts (c'est $S_i^{(l)} \times l!$). On obtient alors

$$E_0(X_t^i) = \frac{1}{2N} - \frac{1}{(2N)^{i+1}} \sum_{l=1}^{\min(i, 2N)} \left[(2N)_{[l]} S_i^{(l)} Pr_0(T_l > t) \right],$$

d'où

$$\begin{aligned} \sum_{t=0}^{\infty} E_0(X_t^i - X_t^{i+1}) &= \sum_{t=0}^{\infty} \left(\frac{1}{(2N)^{i+2}} \sum_{l=1}^{\min(i+1, 2N)} \left[(2N)_{[l]} S_{i+1}^{(l)} Pr_0(T_l > t) \right] \right. \\ &\quad \left. - \frac{1}{(2N)^{i+1}} \sum_{l=1}^{\min(i, 2N)} \left[(2N)_{[l]} S_i^{(l)} Pr_0(T_l > t) \right] \right), \\ &= \frac{1}{(2N)^{i+2}} \sum_{l=1}^{\min(i+1, 2N)} \left[(2N)_{[l]} S_{i+1}^{(l)} \sum_{t=0}^{\infty} Pr_0(T_l > t) \right] \\ &\quad - \frac{1}{(2N)^{i+1}} \sum_{l=1}^{\min(i, 2N)} \left[(2N)_{[l]} S_i^{(l)} \sum_{t=0}^{\infty} Pr_0(T_l > t) \right]. \end{aligned}$$

Puisque

$$\begin{aligned}
& \sum_{t=0}^{\infty} Pr_0(T_l > t) \\
&= Pr_0(T_l > 0) + Pr_0(T_l > 1) + \dots + Pr_0(T_l > k) + \dots, \\
&= \left[Pr_0(T_l = 1) + Pr_0(T_l = 2) + \dots \right] + \left[Pr_0(T_l = 2) + \dots \right] + \dots \\
&\quad \dots + \left[Pr_0(T_l = k + 1) + \dots \right] + \dots, \\
&= \left(1 \times Pr_0(T_l = 1) \right) + \left(2 \times Pr_0(T_l = 2) \right) + \dots + \left(k \times Pr_0(T_l = k) \right) + \dots, \\
&= E_0(T_l),
\end{aligned}$$

et que $E_0(T_1) = 0$, on obtient finalement

$$\sum_{t=0}^{\infty} E_0(X_t^i - X_t^{i+1}) = \sum_{l=2}^{\min(i+1, 2N)} \left((2N)_{[l]} \left[\frac{S_{i+1}^{(l)}}{(2N)^{i+2}} - \frac{S_i^{(l)}}{(2N)^{i+1}} \right] E_0(T_l) \right).$$

On revient à l'équation (2.3.6). On peut maintenant évaluer les deux sommes. On a d'abord, en supposant $2N > 2$,

$$\begin{aligned}
\sum_{t=0}^{\infty} E_0(X_t - X_t^2) &= \sum_{l=2}^{\min(2, 2N)} \left((2N)_{[l]} \left[\frac{S_2^{(l)}}{(2N)^3} - \frac{S_1^{(l)}}{(2N)^2} \right] E_0(T_l) \right), \\
&= (2N)(2N - 1) \left[\frac{S_2^{(2)}}{(2N)^3} - \frac{S_1^{(2)}}{(2N)^2} \right] E_0(T_2).
\end{aligned}$$

Par l'équation (2.3.7), on déduit que $S_n^{(n+1)} = 0$ pour $n \geq 0$, donc $S_1^{(2)} = 0$, et que

$$\begin{aligned}
S_2^{(2)} &= \frac{1}{2!} \sum_{\substack{r_1, r_2 \geq 1 \\ r_1 + r_2 = 2}} \binom{2}{r_1, r_2}, \\
&= \frac{1}{2!} \binom{2}{1, 1}, \\
&= 1,
\end{aligned}$$

d'où

$$\sum_{t=0}^{\infty} E_0(X_t - X_t^2) = \frac{2N-1}{(2N)^2} E_0(T_2).$$

L'espérance de T_2 lorsqu'il n'y a pas de sélection peut facilement être calculé de manière exacte pour notre modèle de population. Dans la section 1.2, on a vu que pour une population diploïde de taille N ($2N$ gènes) suivant le modèle neutre de Wright-Fisher, la probabilité que deux gènes en particulier coalescent en une génération est $1/2N$. On a donc

$$\begin{aligned}
E_0(T_2) &= \sum_{t=1}^{\infty} t \times Pr_0(\text{deux gènes en particulier coalescent en } t \text{ générations}), \\
&= \sum_{t=1}^{\infty} t \times \left(1 - \frac{1}{2N}\right)^{t-1} \left(\frac{1}{2N}\right), \\
&= \frac{1}{2N} \sum_{t=1}^{\infty} t \times \left(1 - \frac{1}{2N}\right)^{t-1}, \\
&= \frac{1}{2N} \times \frac{1}{(1 - [1 - 1/(2N)])^2}, \\
&= 2N.
\end{aligned}$$

Le calcul exact des espérances de T_k lorsque $s = 0$ n'est pas très compliqué si $k = 2$ ou 3. Cependant, dès que $k = 4$, ce calcul devient trop difficile et il vaut mieux

trouver une approximation pour ces espérances. En ayant trouvé l'espérance de T_2 , on peut maintenant conclure que

$$\begin{aligned} \sum_{t=0}^{\infty} E_0(X_t - X_t^2) &= \frac{2N-1}{(2N)^2} 2N , \\ &= 1 - \frac{1}{2N} . \end{aligned}$$

On peut maintenant s'attaquer à la deuxième somme de l'équation (2.3.6). On a, en supposant $2N > 3$,

$$\begin{aligned} \sum_{t=0}^{\infty} E_0(X_t^2 - X_t^3) &= \sum_{l=2}^{\min(3, 2N)} \left((2N)_{[l]} \left[\frac{S_3^{(l)}}{(2N)^4} - \frac{S_2^{(l)}}{(2N)^3} \right] E_0(T_l) \right) , \\ &= (2N)_{[2]} \left[\frac{S_3^{(2)}}{(2N)^4} - \frac{S_2^{(2)}}{(2N)^3} \right] E_0(T_2) + (2N)_{[3]} \left[\frac{S_3^{(3)}}{(2N)^4} - \frac{S_2^{(3)}}{(2N)^3} \right] E_0(T_3) . \end{aligned}$$

On sait que $S_2^{(2)} = 1$, $E_0(T_2) = 2N$ et que $S_2^{(3)} = 0$. De plus, on a

$$\begin{aligned} S_3^{(2)} &= \frac{1}{2!} \sum_{\substack{r_1, r_2 \geq 1 \\ r_1 + r_2 = 3}} \binom{3}{r_1, r_2} , \\ &= \frac{1}{2!} \left(\frac{3!}{2!1!} + \frac{3!}{1!2!} \right) , \\ &= 3 , \end{aligned}$$

et

$$\begin{aligned}
S_3^{(3)} &= \frac{1}{3!} \sum_{\substack{r_1, r_2, r_3 \geq 1 \\ r_1 + r_2 + r_3 = 3}} \binom{3}{r_1, r_2, r_3} , \\
&= \frac{1}{3!} \left(\frac{3!}{1!1!1!} \right) , \\
&= 1 .
\end{aligned}$$

Le calcul exact de l'espérance de T_3 lorsque $s = 0$ est fait de manière analogue à celui de l'espérance de T_2 . Ce calcul, étant fastidieux, est présenté à l'annexe A. On obtient

$$E_0(T_3) = N \left(\frac{8N-3}{3N-1} \right) ,$$

d'où

$$\begin{aligned}
&\sum_{t=0}^{\infty} E_0(X_t^2 - X_t^3) \\
&= 2N(2N-1) \left[\frac{3}{(2N)^4} - \frac{1}{(2N)^3} \right] 2N \\
&\quad + (2N)(2N-1)(2N-2) \left[\frac{1}{(2N)^4} - 0 \right] N \left(\frac{8N-3}{3N-1} \right) , \\
&= \frac{1-2N}{2-6N} .
\end{aligned}$$

On peut maintenant évaluer l'équation (2.3.6) et on trouve

$$\begin{aligned}
F'_A(0) &= h \left(1 - \frac{1}{2N} \right) + (1 - 2h) \left(\frac{1 - 2N}{2 - 6N} \right) , \\
&= \frac{-4 (1 + h) N^2 + 6hN + 2N - 2h}{4N(1 - 3N)} ,
\end{aligned}$$

et donc, de l'équation (2.3.1), on obtient

$$F_A(s) = \frac{1}{2N} + s \left(\frac{-4 (1 + h) N^2 + 6hN + 2N - 2h}{4N(1 - 3N)} \right) + o(s) . \quad (2.3.8)$$

Pour comparer l'approche directe pour la probabilité de fixation par la théorie de la coalescence avec l'approximation par le processus de diffusion (section 2.2), on doit évaluer l'équation (2.2.10) lorsque $u_{AA} = 1$, $u_{AB} = h$, $u_{BB} = 0$, $s = d/2N$ et $x = 1/(2N)$, et on doit la comparer avec l'équation (2.3.8) lorsque N est grand. Pour l'approximation par le processus de diffusion, on obtient

$$F_A \left(\frac{1}{2N} , s \right) = \frac{1}{2N} + s \left(\frac{1 + h}{3} \right) + o(s) ,$$

et, par l'approche directe, on trouve

$$F_A \left(\frac{1}{2N} , s \right) = \frac{1}{2N} + s \left(\frac{1 + h}{3} \right) + o(s) .$$

On obtient donc le même résultat avec les deux méthodes lorsque la taille N de la population est grande.

Chapitre 3

PROBABILITÉ DE FIXATION DANS UN MODÈLE DE SÉLECTION AVEC INTÉRACTION ET CONSANGUINITÉ

3.1. THÉORIE DYNAMIQUE DES JEUX

La théorie dynamique des jeux ([VM]) a beaucoup d'applications dans la vie de tous les jours ; elle peut même nous aider à prendre certaines décisions ! Plus près de ce qui nous intéresse, elle nous permet aussi de comprendre l'évolution d'une population. Puisqu'on se servira de la théorie des jeux dans la section 3.2.1, voici un exemple tiré de [HS] qui permet de l'introduire tout en montrant son utilité.

On suppose qu'il y a deux comportements possibles lors d'une compétition pour l'accouplement d'une espèce en particulier. Le premier comportement consiste à se montrer de plus en plus agressif jusqu'à ce qu'il y ait un blessé ou jusqu'à ce que l'adversaire abandonne et le deuxième comportement consiste seulement à s'afficher et à abandonner si l'adversaire devient plus agressif. On appelle "fauc" le premier comportement et "colombe" le second. On suppose qu'il n'y a pas plus que deux prétendants qui s'affrontent en même temps. Le gagnant peut s'accoupler, et par conséquent il augmente sa valeur sélective, disons de G , alors qu'une blessure réduit la valeur sélective, disons de C .

Lorsque deux individus “colombe” se rencontrent, ils s’affichent et s’intimident jusqu’à ce qu’il y en ait un qui parte (les deux individus ont la même probabilité $1/2$ de gagner). Le gagnant augmente sa valeur sélective de G , alors que celle du perdant ne change pas, car il n’est pas blessé. Le changement moyen de la valeur sélective d’un individu lors d’un tel affrontement est donc $G/2$. Lorsqu’un individu “colombe” rencontre un individu “faucon”, le dernier a raison du premier à coup sûr, car l’individu “colombe” s’enfuit dès que l’individu “faucon” devient trop agressif. La valeur sélective de l’individu “faucon” augmente de G , alors que celle de l’individu “colombe” ne change pas. Enfin, lorsqu’il y a affrontement entre deux individus “faucon”, il y a une bagarre et nécessairement un blessé (les deux individus ont la même probabilité $1/2$ de se blesser). Le gagnant (celui qui n’est pas blessé) a une valeur sélective augmentée de G , alors que le perdant (celui qui est blessé) a une valeur sélective diminuée de C . Le gain moyen en valeur sélective par un individu “faucon” lors d’un tel affrontement est donc $(G - C)/2$, qui est négatif la majorité du temps, car une blessure entraîne souvent un effet négatif plus fort que le bénéfice de l’accouplement. On peut représenter l’ensemble des changements moyens de valeur sélective par la matrice de paiement de la figure 3.1. Il est à spécifier que ce sont les changements moyens de valeur sélective associés aux individus à la gauche de la figure 3.1.

On peut constater que l’avantage d’un type de comportement dépend de la composition de la population. Si la population est composée presque exclusivement d’individus “colombe” (colonne “colombe” du tableau 3.1), alors un individu “faucon” est très avantage, car presque tous ses affrontements se solderont par une victoire sans combat. En effet, on voit dans la matrice de paiement que les individus “faucon” augmenteront leur valeur sélective de G lors de presque tous les affrontements, alors que les individus “colombe” augmenteront la leur de seulement $G/2$. Cependant, si la population est composée majoritairement d’individus “faucon” (colonne “faucon” du tableau 3.1), alors les individus “colombe” sont avantage, car ils évitent les combats et les blessures, ce qui n’est pas le cas des individus “faucon”. On constate dans la matrice de paiement que les individus

	faucon	colombe
faucon	$\frac{(G - C)}{2}$	G
colombe	0	$\frac{G}{2}$

FIG. 3.1. La matrice de paiement

“faucon” gagnent $(G - C)/2$ (on devrait dire perdent $|G - C|/2$ si $C > G$) en valeur sélective, alors que celle des individus “colombe” ne change pas.

Dans une population infinie avec une fréquence x d’individus “faucon” et une fréquence $(1 - x)$ d’individus “colombe”, et des interactions aléatoires entre individus, le changement moyen de valeur sélective sera

$$x\left(\frac{G - C}{2}\right) + (1 - x)G$$

pour les individus “faucon” et

$$(1 - x)\frac{G}{2}$$

pour les individus “colombe”. Comme ces dernières équations sont égales si $x = G/C$, les individus “faucon” sont avantagés si $x < G/C$ (et donc x devrait augmenter par la suite), alors qu’ils sont désavantagés si $x > G/C$ (x devrait diminuer par la suite). La composition de la population devrait se stabiliser et rester en

équilibre lorsque $x = G/C$.

Dans le dernier exemple, chaque individu avait une seule possibilité de comportement et il gardait ce comportement jusqu'à la fin de sa vie. C'était une situation avec des comportements pures ou stratégies pures. Il existe aussi des situations avec des comportements mixtes ou stratégies mixtes. Ce sont des situations dans lesquelles chaque individu opte pour un certain comportement avec une certaine probabilité. En voici un exemple avec une population de deux types d'individus, "bons" et "mauvais", dans laquelle il est possible d'opter pour deux comportements lorsqu'on rencontre un compétiteur : un comportement "faucon" et un comportement "colombe". On a $\mathbf{p}_B = (p_{B1}, p_{B2})$ comme stratégie pour les individus "bons", où p_{B1} est la probabilité de choisir un comportement "faucon" et p_{B2} celle de choisir un comportement "colombe". On a aussi $\mathbf{p}_M = (p_{M1}, p_{M2})$ comme stratégie pour les individus "mauvais", où p_{M1} est la probabilité de choisir un comportement "faucon" et p_{M2} celle de choisir un comportement "colombe". On utilisera les stratégies mixtes à la section 3.2.1.

Une stratégie est dite stable au cours de l'évolution (evolutionarily stable strategy - ESS) si, lorsqu'elle est présente dans toute la population, il n'existe pas d'autre stratégie meilleure pouvant la déloger par la sélection naturelle ([HS]). On associe ici stratégie avec génotype, c'est-à-dire que la stratégie d'un individu est déterminée à partir de son génotype. L'intérêt de l'étude des ESS est que celles-ci aident à mieux comprendre l'évolution dans une population. Pour vérifier si une stratégie est ESS lorsqu'on travaille avec une population infinie, on peut se servir de conditions de stabilité et des valeurs sélectives. Définissons $U_1, U_2, U_3, \dots, U_n$ comme étant n stratégies pures et $w_1, w_2, w_3, \dots, w_n$ les valeurs sélectives associées (rappelons que les valeurs sélectives w_k , $1 \leq k \leq n$, dépendent de la composition de la population). On peut dire qu'une stratégie mixte est une stratégie dans laquelle le choix de $U_1, U_2, U_3, \dots, U_n$ se fait avec les probabilités respectives

p_1, p_2, \dots, p_n . On a que $p_k \geq 0$ et que $\sum_{k=1}^n p_k = 1$. Soit V_n l'ensemble des vecteurs $\mathbf{p} = (p_1, p_2, \dots, p_n)$ de stratégies mixtes possibles et $\bar{\mathbf{p}}$ la stratégie moyenne dans la population. Les w_k dépendent de $\bar{\mathbf{p}}$ car on suppose des interactions aléatoires entre les individus. La valeur sélective d'un individu adoptant une stratégie \mathbf{p} est donc

$$\sum_{k=1}^n p_k w_k(\bar{\mathbf{p}}) = \mathbf{p} \bullet [\mathbf{w}(\bar{\mathbf{p}})] ,$$

où $\mathbf{w}(\bar{\mathbf{p}}) = (w_1(\bar{\mathbf{p}}), w_2(\bar{\mathbf{p}}), w_3(\bar{\mathbf{p}}), \dots, w_n(\bar{\mathbf{p}}))$ et $w_k(\bar{\mathbf{p}})$ est la valeur sélective associée à la stratégie pure k , $1 \leq k \leq n$, lorsque la stratégie moyenne de la population est $\bar{\mathbf{p}}$. Si on a une situation qui consiste en des rencontres aléatoires, on peut supposer que les $w_k(\bar{\mathbf{p}})$ sont linéaires en $\bar{\mathbf{p}}$. On a alors

$$w_i(\bar{\mathbf{p}}) = (M\bar{\mathbf{p}})_i = \sum_{j=1}^n m_{ij} \bar{p}_j ,$$

où $M = (m_{ij})$ est une matrice de paiement de taille n , et m_{ij} est le paiement d'un individu utilisant la stratégie i lors d'une rencontre avec un individu utilisant la stratégie j . On peut donc exprimer la valeur sélective d'un individu utilisant la stratégie \mathbf{p} comme

$$\begin{aligned} \mathbf{p} \bullet M\bar{\mathbf{p}} &= \sum_{i=1}^n p_i w_i(\bar{\mathbf{p}}) , \\ &= \sum_{i,j=1}^n p_i m_{ij} \bar{p}_j . \end{aligned}$$

Supposons maintenant qu'on a une population composée d'une petite fraction ϵ d'individus adoptant la stratégie $\tilde{\mathbf{p}}$ et d'une fraction $1 - \epsilon$ d'individus adoptant la stratégie \mathbf{p} . Alors, la stratégie \mathbf{p} est une stratégie stable au cours de l'évolution si pour tout $\tilde{\mathbf{p}} \in V_n$ tel que $\tilde{\mathbf{p}} \neq \mathbf{p}$, on a

$$\tilde{\mathbf{p}} \bullet M (\epsilon \tilde{\mathbf{p}} + (1 - \epsilon) \mathbf{p}) < \mathbf{p} \bullet M (\epsilon \tilde{\mathbf{p}} + (1 - \epsilon) \mathbf{p}) \quad (3.1.1)$$

pour tout $\epsilon > 0$ plus petit qu'un certain $\epsilon(\tilde{\mathbf{p}})$. En écrivant la dernière équation sous la forme

$$(1 - \epsilon)(\mathbf{p} \bullet M \mathbf{p} - \tilde{\mathbf{p}} \bullet M \mathbf{p}) + \epsilon(\mathbf{p} \bullet M \tilde{\mathbf{p}} - \tilde{\mathbf{p}} \bullet M \tilde{\mathbf{p}}) > 0 ,$$

on trouve que \mathbf{p} est une stratégie stable au cours de l'évolution si et seulement si les conditions suivantes sont satisfaites (voir [HS]) :

(a) la condition d'équilibre

$$\tilde{\mathbf{p}} \bullet M \mathbf{p} \leq \mathbf{p} \bullet M \mathbf{p} , \quad \text{pour tout } \tilde{\mathbf{p}} \text{ dans } V_n ; \quad (3.1.2)$$

(b) la condition de stabilité

$$\tilde{\mathbf{p}} \bullet M \tilde{\mathbf{p}} < \mathbf{p} \bullet M \tilde{\mathbf{p}} \text{ si } \tilde{\mathbf{p}} \neq \mathbf{p} \text{ et } \tilde{\mathbf{p}} \bullet M \mathbf{p} = \mathbf{p} \bullet M \mathbf{p} . \quad (3.1.3)$$

La condition d'équilibre signifie que \mathbf{p} doit être la meilleure stratégie à adopter fasse à un individu dont la stratégie est \mathbf{p} . La condition de stabilité est cependant nécessaire, car il peut y avoir d'autres stratégies aussi efficace que la stratégie \mathbf{p} lors d'une telle rencontre. La condition de stabilité signifie que dans les cas où il y a égalité dans la condition d'équilibre, avec une stratégie $\tilde{\mathbf{p}}$ par exemple, la stratégie \mathbf{p} doit être la meilleure stratégie à adopter, et ce, sans qu'elle n'ait d'égale, fasse à un individu dont la stratégie est $\tilde{\mathbf{p}}$. Ces conditions nous amènent au théorème suivant :

Théorème 3.1.1. *Une stratégie \mathbf{p} est une ESS si et seulement si*

$$\mathbf{p} \bullet M \mathbf{r} > \mathbf{r} \bullet M \mathbf{r} ,$$

pour tout $\mathbf{r} \neq \mathbf{p}$ dans un voisinage de \mathbf{p} dans V_n .

Démonstration. (voir [HS])

Supposons que \mathbf{p} est une ESS. Comme on veut \mathbf{r} dans un voisinage de \mathbf{p} , on peut exprimer \mathbf{r} ainsi : $\mathbf{r} = \epsilon \tilde{\mathbf{p}} + (1 - \epsilon)\mathbf{p}$ pour un certain $\tilde{\mathbf{p}}$ dans V_n et un certain $\epsilon > 0$ assez petit. On peut même prendre la stratégie $\tilde{\mathbf{p}}$ telle que $\tilde{\mathbf{p}}$ appartient à l'ensemble compact $C = \{\mathbf{r} \in V_n | r_j = 0 \text{ pour } j \in \text{supp}(\mathbf{p})\}$, où $\text{supp}(\mathbf{p})$ est l'ensemble des indices k tels que $p_k > 0$, qui est l'ensemble des "faces" de V_n qui ne contiennent pas \mathbf{p} . On a donc que pour tout $\tilde{\mathbf{p}} \in C$, l'inégalité (3.1.1) s'applique pour $0 < \epsilon < \epsilon(\tilde{\mathbf{p}})$. On peut choisir $\epsilon(\tilde{\mathbf{p}})$ de manière à ce qu'il soit continu et alors poser $0 < \epsilon^* = \min\{\epsilon(\tilde{\mathbf{p}}) | \tilde{\mathbf{p}} \in C\}$. On a donc que l'inégalité (3.1.1) s'applique pour tout $0 < \epsilon < \epsilon^*$ et alors

$$\tilde{\mathbf{p}} \bullet M (\epsilon \tilde{\mathbf{p}} + (1 - \epsilon)\mathbf{p}) < \mathbf{p} \bullet M (\epsilon \tilde{\mathbf{p}} + (1 - \epsilon)\mathbf{p}) ,$$

ce qui entraîne

$$\epsilon [\tilde{\mathbf{p}} \bullet M (\epsilon \tilde{\mathbf{p}} + (1 - \epsilon)\mathbf{p})] < \epsilon [\mathbf{p} \bullet M (\epsilon \tilde{\mathbf{p}} + (1 - \epsilon)\mathbf{p})] .$$

En ajoutant $(1 - \epsilon)\mathbf{p} \bullet M (\epsilon \tilde{\mathbf{p}} + (1 - \epsilon)\mathbf{p})$ des deux côtés de l'inégalité ci-dessus et en utilisant $\mathbf{r} = \epsilon \tilde{\mathbf{p}} + (1 - \epsilon)\mathbf{p}$, on trouve finalement l'inégalité

$$\mathbf{r} \bullet M \mathbf{r} < \mathbf{p} \bullet M \mathbf{r}$$

pour tout $\mathbf{r} \neq \mathbf{p}$ dans un voisinage de \mathbf{p} . Inversement, supposons que

$$\mathbf{r} \bullet M \mathbf{r} < \mathbf{p} \bullet M \mathbf{r}$$

pour tout $\mathbf{r} \neq \mathbf{p}$ dans un voisinage de \mathbf{p} . Pour tout $\tilde{\mathbf{p}} \in V_n$ tel que $\tilde{\mathbf{p}} \neq \mathbf{p}$, il existe un $\epsilon(\tilde{\mathbf{p}})$ tel que

$$\mathbf{r} = \epsilon \tilde{\mathbf{p}} + (1 - \epsilon) \mathbf{p}$$

est dans le voisinage de \mathbf{p} pour tout $0 < \epsilon < \epsilon(\tilde{\mathbf{p}})$. On a alors

$$[\epsilon \tilde{\mathbf{p}} + (1 - \epsilon) \mathbf{p}] \bullet M[\epsilon \tilde{\mathbf{p}} + (1 - \epsilon) \mathbf{p}] < \mathbf{p} \bullet M[\epsilon \tilde{\mathbf{p}} + (1 - \epsilon) \mathbf{p}] .$$

On soustrait ensuite $(1 - \epsilon) \mathbf{p} \bullet M[\epsilon \tilde{\mathbf{p}} + (1 - \epsilon) \mathbf{p}]$ des deux côtés de l'inégalité pour obtenir

$$\epsilon \tilde{\mathbf{p}} \bullet M[\epsilon \tilde{\mathbf{p}} + (1 - \epsilon) \mathbf{p}] < \epsilon \mathbf{p} \bullet M[\epsilon \tilde{\mathbf{p}} + (1 - \epsilon) \mathbf{p}] ,$$

ce qui est équivalent à

$$\tilde{\mathbf{p}} \bullet M[\epsilon \tilde{\mathbf{p}} + (1 - \epsilon) \mathbf{p}] < \mathbf{p} \bullet M[\epsilon \tilde{\mathbf{p}} + (1 - \epsilon) \mathbf{p}] .$$

Cette inégalité étant satisfaite pour tout $\tilde{\mathbf{p}} \in V_n$ tel que $\tilde{\mathbf{p}} \neq \mathbf{p}$ et pour tout $0 < \epsilon < \epsilon(\tilde{\mathbf{p}})$, on en conclut que \mathbf{p} est une ESS.

□

On a déjà vu à la section 2.3 que chaque allèle, dans une population finie, se fixe ou disparaît à long terme. Puisqu'un génotype est associé à une stratégie, la probabilité de fixation est aussi la probabilité qu'une stratégie se fixe dans la population. On peut donc utiliser la probabilité de fixation pour trouver des stratégies optimales dans des populations finies. Une probabilité de fixation, d'un allèle particulier A , plus grande que sa fréquence initiale est synonyme d'avantage

sélectif de cet allèle. On peut alors se demander quelle stratégie a un avantage sélectif sur toute autre et si cette stratégie correspond à un ESS telle que définie précédemment.

On utilisera les matrices de jeux avec les stratégies mixtes dans le chapitre 3.2.1. On considérera une population dans laquelle il y aura deux stratégies possibles, 1 et 2. On aura trois types d'individus selon le génotype (AA , AB ou BB), et chacun aura un vecteur de probabilité différent qui contiendra la probabilité de choisir la première stratégie et celle de choisir la seconde. Ces vecteurs représenteront donc trois stratégies mixtes différentes, qui seront associées aux trois génotypes. On aura aussi une matrice de paiement $M_{2 \times 2}$ dont chaque élément M_{ij} représentera un paiement pour l'utilisation de la stratégie i lors d'une rencontre avec un autre individu utilisant la stratégie j .

3.2. PROBABILITÉ DE FIXATION DANS UNE POPULATION AVEC AUTOFÉCONDATION PARTIELLE

3.2.1. Probabilité de fixation dans une population avec autofécondation partielle et interaction

On suppose une population de N individus diploïdes (donc $2N$ gènes), dans laquelle les générations sont discrètes et ne se chevauchent pas. À chaque génération, il y a N individus. On observe un locus où il y a deux allèles : A et B . On s'intéresse à la probabilité de fixation de l'allèle A sachant que sa fréquence initiale dans la population est $1/2N$. Soit α la probabilité qu'un individu se reproduise avec lui-même (autofécondation) et donc $1-\alpha$ celle qu'il se reproduise avec un autre individu au hasard. Cela est possible et fréquent dans les populations hermaphrodites. On suppose qu'il y a interaction entre les jeunes ainsi produits et qu'il y a deux stratégies possibles, disons 1 et 2, déterminées au locus considéré. Les valeurs sélectives associées aux trois différents génotypes sont alors les suivantes :

$$\begin{aligned}
w_{AA}(t) &= 1 + s \mathbf{p}_{AA} \bullet M\bar{\mathbf{p}}(t) , \\
w_{AB}(t) &= 1 + s \mathbf{p}_{AB} \bullet M\bar{\mathbf{p}}(t) , \\
w_{BB}(t) &= 1 + s \mathbf{p}_{BB} \bullet M\bar{\mathbf{p}}(t) ,
\end{aligned} \tag{3.2.1}$$

où s représente l'intensité de la sélection, M est une matrice de jeu contenant les paiements lors de rencontres entre différents stratégestes (voir la section 3.1), et $\mathbf{p}_{ij} = (p_{ij}, 1 - p_{ij})$, où p_{ij} est la probabilité qu'un individu de génotype ij choisisse la stratégie 1. La stratégie moyenne parmi l'infinité de jeunes produits par les adultes de la génération t , avant la sélection est représentée par

$$\bar{\mathbf{p}}(t) = (X(t))^2 \mathbf{p}_{AA} + 2X(t)(1 - X(t)) \mathbf{p}_{AB} + (1 - X(t))^2 \mathbf{p}_{BB} ,$$

où $X(t)$ est la fréquence de l'allèle A à la génération t . Pour simplifier les calculs, on suppose que le modèle est additif, c'est-à-dire

$$\mathbf{p}_{AB} = \frac{1}{2} \mathbf{p}_{AA} + \frac{1}{2} \mathbf{p}_{BB},$$

et dans ce cas

$$\bar{\mathbf{p}}(t) = (\mathbf{p}_{AA} - \mathbf{p}_{BB})X(t) + \mathbf{p}_{BB}.$$

Le cycle d'une génération à la suivante, avec entre parenthèses la notation utilisée pour les fréquences des génotypes ij à chaque étape, est résumé dans le tableau suivant :

$$\begin{array}{c}
N \text{ adultes de la génération } t \ (Q_{ij}(t)) \xrightarrow{\text{reproduction}} \text{infinité de jeunes } (Q'_{ij}(t)) \xrightarrow{\text{sélection}} \\
\text{infinité d'adultes } (\bar{Q}_{ij}(t)) \xrightarrow{\text{tirage}} N \text{ adultes de la génération } t+1 \ (Q_{ij}(t+1)).
\end{array}$$

L'effet de la reproduction sur les fréquences des génotypes est :

$$\begin{aligned}
Q'_{AA}(t) &= \alpha \left(Q_{AA}(t) + \frac{1}{4} Q_{AB}(t) \right) + (1 - \alpha) (X^2(t)) , \\
Q'_{AB}(t) &= \alpha \left(\frac{1}{2} Q_{AB}(t) \right) + (1 - \alpha) (2 X(t)(1 - X(t))) , \\
Q'_{BB}(t) &= \alpha \left(Q_{BB}(t) + \frac{1}{4} Q_{AB}(t) \right) + (1 - \alpha) (1 - X(t))^2 .
\end{aligned} \tag{3.2.2}$$

Notons que $X(t) = Q_{AA}(t) + \frac{1}{2} Q_{AB}(t)$.

Après la reproduction, il y a sélection entre les jeunes, ce qui transforme les fréquences des génotypes comme suit :

$$\begin{aligned}
\tilde{Q}_{AA}(t) &= \frac{w_{AA}(t) \times Q'_{AA}(t)}{(w_{AA}(t) \times Q'_{AA}(t)) + (w_{AB}(t) \times Q'_{AB}(t)) + (w_{BB}(t) \times Q'_{BB}(t))} , \\
\tilde{Q}_{AB}(t) &= \frac{w_{AB}(t) \times Q'_{AB}(t)}{(w_{AA}(t) \times Q'_{AA}(t)) + (w_{AB}(t) \times Q'_{AB}(t)) + (w_{BB}(t) \times Q'_{BB}(t))} , \\
\tilde{Q}_{BB}(t) &= \frac{w_{BB}(t) \times Q'_{BB}(t)}{(w_{AA}(t) \times Q'_{AA}(t)) + (w_{AB}(t) \times Q'_{AB}(t)) + (w_{BB}(t) \times Q'_{BB}(t))} .
\end{aligned} \tag{3.2.3}$$

Ensuite, on tire N individus qui seront les adultes de la génération $t + 1$. Comme le tirage se fait au hasard, on a

$$E(Q_{ij}(t+1) | Q_{AA}(t), Q_{AB}(t), Q_{BB}(t)) = E(\tilde{Q}_{ij}(t) | Q_{AA}(t), Q_{AB}(t), Q_{BB}(t)). \tag{3.2.4}$$

Ce qui nous intéresse est la probabilité de fixation de A et on a vu à la section 2.3 qu'on peut l'exprimer comme suit (voir l'équation (2.3.1)) :

$$F_A(s) = F_A(0) + sF'_A(0) + o(s) = \frac{1}{2N} + sF'_A(0) + o(s) .$$

Il ne nous reste donc qu'à trouver $F'_A(0)$. À la section 2.3, on passait par le calcul exact de l'espérance des T_k (lorsque $s = 0$) pour obtenir $F'_A(0)$, où T_k représentait le nombre de générations qu'il faut remonter avant d'arriver au MRCA à partir de k gènes à la génération t . On avait alors un modèle de population sans autofécondation partielle ni interaction et tel que le calcul exact de l'espérance des T_k n'était pas très difficile à calculer, du moins lorsque k était petit. Ici, puisqu'on étudie un modèle avec autofécondation partielle et interaction, le calcul exact de l'espérance des T_k serait plus compliqué à obtenir. On remplacera donc le calcul exact de l'espérance par une approximation de celle-ci en supposant que N est grand. Reprenons donc à partir du commencement la démarche de la section 2.3 proposée initialement dans [LL] pour trouver $F'_A(0)$. Comme on a vu plus tôt, à long terme soit A se fixe (i.e., la fréquence de A est $2N/2N = 1$), soit il disparaît de la population (i.e., la fréquence de A est $0/2N = 0$). En fait, la variable $X(t)$ converge en moyenne vers la variable

$$X(\infty) = \begin{cases} 1 & \text{avec probabilité } F_A(s), \\ 0 & \text{avec probabilité } 1 - F_A(s), \end{cases}$$

lorsque $t \rightarrow \infty$ (voir la deuxième démonstration du chapitre 2.3) et

$$E_s(X(\infty)) - E_s(X(0)) = \sum_{t=0}^{\infty} E_s[X(t+1) - X(t)] .$$

Puisque $E_s(X(\infty)) = F_A(s)$ et $E_s(X(0)) = x_0 = 1/2N$, on a

$$F_A(s) = \frac{1}{2N} + \sum_{t=0}^{\infty} E_s[X(t+1) - X(t)] .$$

En conditionnant sur les fréquences des génotypes à la génération précédente, on se retrouve finalement avec l'expression suivante pour la probabilité de fixation :

$$F_A(s) = \frac{1}{2N} + \sum_{t=0}^{\infty} \sum_{\mathbf{q}} E_s(X(t+1) - X(t) | \mathbf{Q}(t) = \mathbf{q}) \times Pr_s(\mathbf{Q}(t) = \mathbf{q}) ,$$

où $\mathbf{Q}(t) = \{Q_{AA}(t), Q_{AB}(t), Q_{BB}(t)\}$ et $\mathbf{q} = \{q_{AA}, q_{AB}, q_{BB}\}$. Notons qu'on a alors $X(t) = x = q_{AA} + q_{AB}/2$. En dérivant (dans des conditions de régularité, la dérivée de la somme est la somme des dérivées, voir [LL]) et en évaluant à $s = 0$, on obtient que

$$\begin{aligned} F'_A(0) &= \sum_{t=0}^{\infty} \sum_{\mathbf{q}} \left[\left(\frac{d}{ds} E_s(X(t+1) - X(t) | \mathbf{Q}(t) = \mathbf{q}) \right) \times Pr_s(\mathbf{Q}(t) = \mathbf{q}) \right. \\ &\quad \left. + E_s(X(t+1) - X(t) | \mathbf{Q}(t) = \mathbf{q}) \times \left(\frac{d}{ds} Pr_s(\mathbf{Q}(t) = \mathbf{q}) \right) \right]_{s=0} , \\ &= \sum_{t=0}^{\infty} \sum_{\mathbf{q}} \left[\left(\frac{d}{ds} E_s(X(t+1) - X(t) | \mathbf{Q}(t) = \mathbf{q}) \right) \right]_{s=0} \times Pr_0(\mathbf{Q}(t) = \mathbf{q}) , \end{aligned} \quad (3.2.5)$$

où

$$Pr_0(\mathbf{Q}(t) = \mathbf{q}) = Pr(\mathbf{Q}(t) = \mathbf{q})|_{s=0} ,$$

et

$$\left[E_s(X(t+1) - X(t) | \mathbf{Q}(t) = \mathbf{q}) \times \frac{d}{ds} (Pr_s(\mathbf{Q}(t) = \mathbf{q})) \right]_{s=0} = 0 ,$$

car, sans sélection ($w_{AA} = w_{AB} = w_{BB} = 1$), la fréquence de A ne change pas puisque

$$\begin{aligned}
E_s \left(X(t+1) | \mathbf{Q}(t) = \mathbf{q} \right) \Big|_{s=0} &= E_0 \left(\tilde{Q}_{AA}(t) + \frac{1}{2} \tilde{Q}_{AB}(t) | \mathbf{Q}(t) = \mathbf{q} \right) , \\
&= E_0 \left(Q'_{AA}(t) + \frac{1}{2} Q'_{AB}(t) | \mathbf{Q}(t) = \mathbf{q} \right) , \\
&= E_0 \left[\alpha \left(Q_{AA}(t) + \frac{1}{4} Q_{AB}(t) \right) + (1 - \alpha) \left(X^2(t) \right) \right. \\
&\quad \left. + \frac{\alpha}{2} \left(\frac{1}{2} Q_{AB}(t) \right) + \frac{1 - \alpha}{2} \left(2 X(t) [1 - X(t)] \right) \right] \Big| \mathbf{Q}(t) = \mathbf{q} , \\
&= \alpha x + (1 - \alpha) (x^2 + x(1 - x)) , \\
&= x .
\end{aligned}$$

Il faut noter qu'en présence de sélection, l'espérance de $(X(t+1) - X(t) | \mathbf{Q}(t) = \mathbf{q})$ dans (3.2.5) est différente selon le modèle utilisé. Pour notre modèle de population avec autofécondation partielle et interaction entre les individus, en utilisant les équations (3.2.4), (3.2.3), (3.2.2) et (3.2.1), on trouve

$$E(X(t+1) - X(t) | \mathbf{Q}(t) = \mathbf{q}) = (a/b) - x ,$$

où

$$\begin{aligned}
a = x + \alpha &\left[\left(q_{AA} + \frac{3}{8} q_{AB} \right) \left(x s_{\mathbf{P}_{AA}} \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB}) + s_{\mathbf{P}_{AA}} \bullet M \mathbf{p}_{BB} \right) \right. \\
&\quad \left. + \left(\frac{1}{8} q_{AB} \right) \left(x s_{\mathbf{P}_{BB}} \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB}) + s_{\mathbf{P}_{BB}} \bullet M \mathbf{p}_{BB} \right) \right] \\
&+ (1 - \alpha) \left[\left(x^2 + \frac{x(1-x)}{2} \right) \left(x s_{\mathbf{P}_{AA}} \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB}) + s_{\mathbf{P}_{AA}} \bullet M \mathbf{p}_{BB} \right) \right. \\
&\quad \left. + \left(\frac{x(1-x)}{2} \right) \left(x s_{\mathbf{P}_{BB}} \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB}) + s_{\mathbf{P}_{BB}} \bullet M \mathbf{p}_{BB} \right) \right]
\end{aligned}$$

et

$$b = 1 + x^2 s \mathbf{p}_{AA} \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB}) + x s \mathbf{p}_{AA} \bullet M \mathbf{p}_{BB} \\ + x(1-x) s \mathbf{p}_{BB} \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB}) + (1-x) s \mathbf{p}_{BB} \bullet M \mathbf{p}_{BB} .$$

L'expression a/b est de la forme $\frac{a}{1-r}$, où $|r| < 1$ pour s suffisamment petit, et alors on peut utiliser le fait que $\frac{a}{1-r} = a \sum_{i=0}^{\infty} r^i$. On trouve alors

$$E_s(X(t+1) - X(t) | \mathbf{Q}(t) = \mathbf{q}) \\ = s[a_1 \mathbf{p}_{AA} \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB}) + b_1 \mathbf{p}_{AA} \bullet M \mathbf{p}_{BB} \\ + c_1 \mathbf{p}_{BB} \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB}) + d_1 \mathbf{p}_{BB} \bullet M \mathbf{p}_{BB}] + o(s) ,$$

où $o(s)$ est une fonction petit ordre de s et

$$a_1 = \alpha \left[x \left(q_{AA} + \frac{3}{8} q_{AB} \right) \right] + (1-\alpha) \left[x \left(x^2 + \frac{x(1-x)}{2} \right) \right] - x^3 ,$$

$$b_1 = \alpha \left[q_{AA} + \frac{3}{8} q_{AB} \right] + (1-\alpha) \left[x^2 + \frac{x(1-x)}{2} \right] - x^2 ,$$

$$c_1 = \alpha \left[\frac{x}{8} q_{AB} \right] + (1-\alpha) \left[\frac{x^2(1-x)}{2} \right] - x^2(1-x) ,$$

$$d_1 = \alpha \left[\frac{1}{8} q_{AB} \right] + (1-\alpha) \left[\frac{x(1-x)}{2} \right] - x(1-x) .$$

En dérivant par rapport à s et en évaluant à $s = 0$, on obtient

$$\begin{aligned}
& \frac{d}{ds} E_s(X(t+1) - X(t) | \mathbf{Q}(t) = \mathbf{q}) \Big|_{s=0} \\
&= a_1 \mathbf{p}_{AA} \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB}) + b_1 \mathbf{p}_{AA} \bullet M\mathbf{p}_{BB} \\
&+ c_1 \mathbf{p}_{BB} \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB}) + d_1 \mathbf{p}_{BB} \bullet M\mathbf{p}_{BB} .
\end{aligned}$$

En remplaçant x par $q_{AA} + \frac{1}{2}q_{AB}$, on trouve

$$\begin{aligned}
& \frac{d}{ds} E_s(X(t+1) - X(t) | \mathbf{Q}(t) = \mathbf{q}) \Big|_{s=0} \\
&= a_2 \mathbf{p}_{AA} \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB}) + b_2 \mathbf{p}_{AA} \bullet M\mathbf{p}_{BB} \\
&+ c_2 \mathbf{p}_{BB} \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB}) + d_2 \mathbf{p}_{BB} \bullet M\mathbf{p}_{BB} ,
\end{aligned}$$

où

$$\begin{aligned}
a_2 = & \left(\frac{1}{2} + \frac{\alpha}{2}\right) q_{AA}^2 + \left(\frac{1}{2} + \frac{3\alpha}{8}\right) q_{AA}q_{AB} + \left(\frac{1}{8} + \frac{\alpha}{16}\right) q_{AB}^2 + \left(-\frac{1}{2} - \frac{\alpha}{2}\right) q_{AA}^3 \\
& + \left(-\frac{3}{4} - \frac{3\alpha}{4}\right) q_{AA}^2q_{AB} + \left(-\frac{3}{8} - \frac{3\alpha}{8}\right) q_{AA}q_{AB}^2 + \left(-\frac{1}{16} - \frac{\alpha}{16}\right) q_{AB}^3 ,
\end{aligned}$$

$$\begin{aligned}
b_2 = & \left(\frac{1}{2} + \frac{\alpha}{2}\right) q_{AA} + \left(\frac{1}{4} + \frac{\alpha}{8}\right) q_{AB} + \left(-\frac{1}{2} - \frac{\alpha}{2}\right) q_{AA}^2 + \left(-\frac{1}{2} - \frac{\alpha}{2}\right) q_{AA}q_{AB} \\
& + \left(-\frac{1}{8} - \frac{\alpha}{8}\right) q_{AB}^2 ,
\end{aligned}$$

$$c_2 = \left(-\frac{1}{2} - \frac{\alpha}{2}\right) q_{AA}^2 + \left(-\frac{1}{2} - \frac{3\alpha}{8}\right) q_{AA}q_{AB} + \left(-\frac{1}{8} - \frac{\alpha}{16}\right) q_{AB}^2 + \left(\frac{1}{2} + \frac{\alpha}{2}\right) q_{AA}^3 \\ + \left(\frac{3}{4} + \frac{3\alpha}{4}\right) q_{AA}^2q_{AB} + \left(\frac{3}{8} + \frac{3\alpha}{8}\right) q_{AA}q_{AB}^2 + \left(\frac{1}{16} + \frac{\alpha}{16}\right) q_{AB}^3 ,$$

$$d_2 = \left(-\frac{1}{2} - \frac{\alpha}{2}\right) q_{AA} + \left(-\frac{1}{4} - \frac{\alpha}{8}\right) q_{AB} + \left(\frac{1}{2} + \frac{\alpha}{2}\right) q_{AA}^2 + \left(\frac{1}{2} + \frac{\alpha}{2}\right) q_{AA}q_{AB} \\ + \left(\frac{1}{8} + \frac{\alpha}{8}\right) q_{AB}^2 .$$

Finalement, en multipliant par $Pr_0(\mathbf{Q}(t) = \mathbf{q})$ et en sommant sur t et \mathbf{q} , on a

$$F'_A(0) = a_3 \sum_{t \geq 0} E_0(Q_{AA}(t) - Q_{AA}^2(t)) + b_3 \sum_{t \geq 0} E_0(Q_{AB}(t)) + c_3 \sum_{t \geq 0} E_0(Q_{AA}^2(t) - Q_{AA}^3(t)) \\ + d_3 \sum_{t \geq 0} E_0(Q_{AA}(t)Q_{AB}(t)) + e_3 \sum_{t \geq 0} E_0(Q_{AB}^2(t)) + g_3 \sum_{t \geq 0} E_0(Q_{AA}^2(t)Q_{AB}(t)) \\ + h_3 \sum_{t \geq 0} E_0(Q_{AA}(t)Q_{AB}^2(t)) + k_3 \sum_{t \geq 0} E_0(Q_{AB}^3(t)) ,$$

(3.2.6)

où $E_0(Q_{ij}(t))$ est l'espérance de $Q_{ij}(t)$ sous l'hypothèse qu'il n'y a pas de sélection ($s = 0$) et

$$a_3 = \left(\frac{1}{2} + \frac{\alpha}{2} \right) (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M \mathbf{p}_{BB} ,$$

$$b_3 = \left(\frac{1}{4} + \frac{\alpha}{8} \right) (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M \mathbf{p}_{BB} ,$$

$$c_3 = \left(\frac{1}{2} + \frac{\alpha}{2} \right) (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M (\mathbf{p}_{AA} - \mathbf{p}_{BB}) ,$$

$$\begin{aligned} d_3 = & \left(\frac{1}{2} + \frac{3\alpha}{8} \right) (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \\ & + \left(-\frac{1}{2} - \frac{\alpha}{2} \right) (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M \mathbf{p}_{BB} , \end{aligned}$$

$$\begin{aligned} e_3 = & \left(\frac{1}{8} + \frac{\alpha}{16} \right) (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \\ & + \left(-\frac{1}{8} - \frac{\alpha}{8} \right) (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M \mathbf{p}_{BB} , \end{aligned}$$

$$g_3 = \left(-\frac{3}{4} - \frac{3\alpha}{4} \right) (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M (\mathbf{p}_{AA} - \mathbf{p}_{BB}) ,$$

$$h_3 = \left(-\frac{3}{8} - \frac{3\alpha}{8} \right) (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M (\mathbf{p}_{AA} - \mathbf{p}_{BB}) ,$$

$$k_3 = \left(-\frac{1}{16} - \frac{\alpha}{16} \right) (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M (\mathbf{p}_{AA} - \mathbf{p}_{BB}) .$$

Pour calculer les sommes des espérances ci-dessus, on utilise le processus de coalescence. En effet, pour le deuxième terme de l'équation (3.2.6) par exemple, on peut constater que

$$\sum_{t \geq 0} E_0(Q_{AB}(t)) = \sum_{t \geq 0} E_0 \left(\frac{1}{N} \sum_{k=1}^N \eta_k(t) \right) ,$$

où

$$\eta_k(t) = \begin{cases} 1 & \text{si l'individu } k \text{ est } AB \text{ à la génération } t , \\ 0 & \text{sinon.} \end{cases}$$

Comme les $\eta_k(t)$ ont la même distribution pour tout k , on a

$$\begin{aligned} \sum_{t \geq 0} E_0(Q_{AB}(t)) &= \sum_{t \geq 0} \frac{1}{N} \sum_{k=1}^N E_0(\eta_k(t)) = \sum_{t \geq 0} E_0(\eta_k(t)) , \\ &= \sum_{t \geq 0} Pr_0(\text{l'individu } k \text{ soit } AB \text{ à la génération } t) . \end{aligned}$$

On se servira du processus de coalescence car nous suivrons les gènes ancestraux de l'individu k de la génération t à la génération 0 pour connaître la probabilité que l'individu k soit de génotype AB sachant qu'à la génération 0, la fréquence de l'allèle A est $X(0) = \frac{1}{2N}$. On aura besoin de la matrice de transition Π sur l'ensemble de toutes les configurations possibles de gènes ancestraux d'une génération à la précédente en incluant l'éventualité de coalescence. Par la suite nous pourrons considérer Π^t , la t -ième puissance de Π , qui permettra de calculer la probabilité qu'un individu k de la génération t soit de génotype AB . En considérant toutes les espérances à calculer, on aura une matrice Π sur quinze états au total, car il y a au maximum six gènes ancestraux à suivre. En effet, ayant à considérer des espérances de termes de degré trois au maximum dans l'équation (3.2.6), on aura à suivre les gènes ancestraux d'au plus trois individus diploïdes de la génération t , donc d'au plus six gènes. Voici les quinze configurations possibles numérotées de 1 à 15 (les ovales représentent des individus et les petits cercles représentent des gènes, les gènes ancestraux étant en noir) :

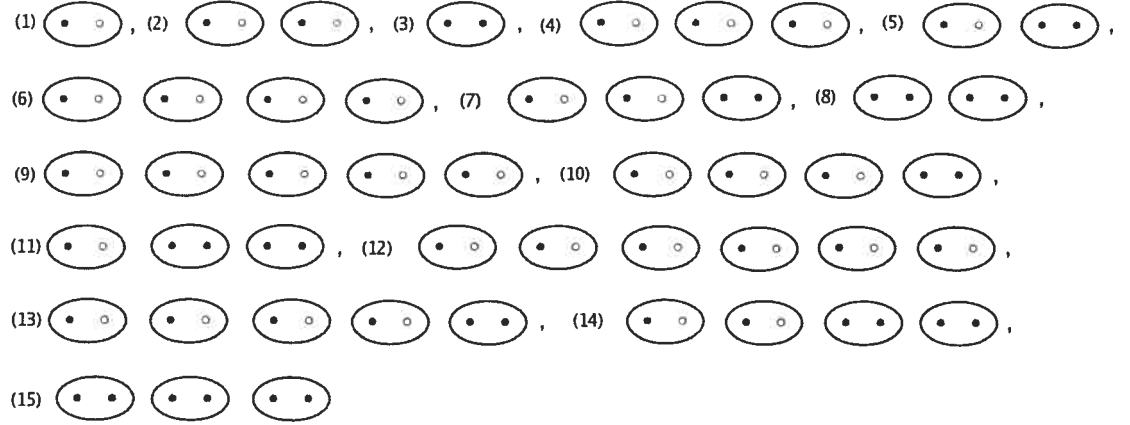


FIG. 3.2. Les quinze états possibles pour la configuration ancestrale de six gènes.

En utilisant la matrice Π et la t -ième itérée, Π^t , on obtient que

$$\sum_{t \geq 0} E_0(Q_{AB}(t)) = \sum_{t \geq 0} \left[(\Pi^t)_{[3,2]} + (\Pi^t)_{[3,3]} \right] \times \frac{1}{2N}, \quad (3.2.7)$$

où $(\Pi^t)_{[3,j]}$ est la probabilité de passer de l'état 3 à la génération t , à l'état j à la génération 0, car, premièrement, en remontant le temps, on doit commencer par suivre les lignées ancestrales des deux gènes d'un individu AB de la génération t (l'état de la matrice Π^t qui représente deux gènes ancestraux dans un individu est l'état 3), et deuxièmement, ces deux gènes étant de type différent (A et B), ils ne peuvent coalescer ensemble. Il doit donc encore y avoir deux gènes ancestraux que l'on suit à la génération 0 (état 2 ou 3). En fait, on verra plus loin qu'en pratique, lorsque N est grand, la seule possibilité d'état pour les gènes ancestraux à la génération 0 est celle où il y a deux gènes qui ne sont pas dans le même individu (état 2 de la matrice Π^t). La figure 3.3 illustre le chemin de ces deux lignées ancestrales. On doit ensuite multiplier par $1/2N$ car c'est la probabilité qu'une des lignées soit de type A , sachant que la fréquence de A à la génération 0 est $1/2N$ (l'autre lignée est alors automatiquement de type B).

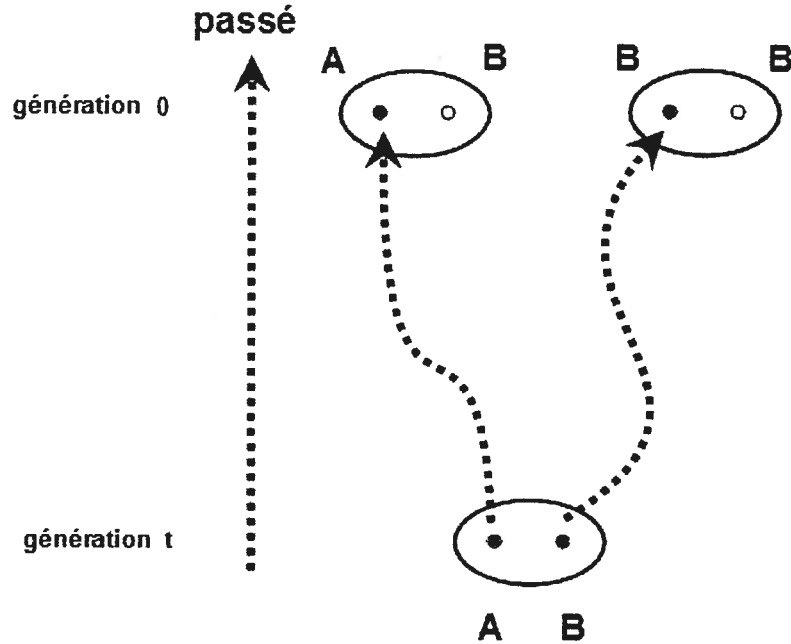


FIG. 3.3. Façon d'avoir le génotype AB à la génération t .

En supposant N grand, nous ferons maintenant appel au lemme C.0.1, cité et expliqué dans l'annexe C. Ce lemme, qui est de Möhle (voir [M]), peut être utilisé, car on peut écrire la matrice Π sous forme $H + c_N R_N$, où les éléments de la matrice Π sont déduits logiquement et c_N est une suite de constantes qui tend vers 0 lorsque N tend vers l'infini ($c_N = 1/2N$ dans notre cas). Les matrices H et R_N sont des matrices de même taille (et avec les mêmes états) que Π et leurs éléments, qui ont été obtenus à partir des éléments de la matrice Π , sont donnés respectivement dans les annexes B.1 et B.3. La matrice H contient la partie des probabilités de transition de la matrice Π qui est associée à l'autofécondation, alors que la matrice R_N contient la partie des probabilités de transition de la matrice Π qui est associée à la reproduction au hasard. Puisque nous utiliserons la deuxième partie du lemme C.0.1, nous avons besoin de l'existence de deux autres matrices, les matrices R et P . La matrice R , où $R = \lim_{N \rightarrow \infty} R_N$, existe car on peut constater, à l'annexe B.3, que la limite, lorsque N tend vers l'infini, des éléments de R_N converge. De plus, puisque $P = \lim_{n \rightarrow \infty} H^n$, on peut constater que P existe et trouver ses éléments à partir de la matrice H . Les éléments des matrices P et R sont donnés respectivement dans les annexes B.2 et B.4 et ont

été obtenus en calculant les limites. On peut donc utiliser la deuxième partie du lemme C.0.1 de Möhle, duquel on tire que

$$\lim_{N \rightarrow \infty} \Pi^{[2N\tau]} = P e^{\tau PRP} \quad \text{pour tout } \tau \geq 0 ,$$

où $[2N\tau]$ désigne la valeur entière de $2N\tau$. Plus d'explications sur ce lemme sont données dans l'annexe C. Les éléments de la matrice $e^{\tau PRP}$ sont donnés dans l'annexe B.5 et ont été obtenus à partir des matrices P et R . En appliquant ce résultat à l'équation (3.2.7), on obtient

$$\lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AB}(t)) = P_{[3,2]} \int_0^\infty (e^{PRP\tau})_{[2,2]} d\tau , \quad (3.2.8)$$

où $P_{[3,2]}$ est l'élément de la troisième ligne et de la deuxième colonne de la matrice P , et $(e^{PRP\tau})_{[2,2]}$ est l'élément de la deuxième ligne et de la deuxième colonne de la matrice $e^{PRP\tau}$. Le terme $(\Pi^t)_{[3,3]}$ de l'équation (3.2.7) disparaît en passant à la limite car la probabilité que les deux lignées se retrouvent dans le même individu à la génération 0 tend vers 0 lorsque N tend vers l'infini. L'équation (3.2.8) est due au fait qu'on peut représenter la somme sur $t \geq 0$ de $(\Pi^t)_{[i,j]} \times \frac{1}{2N}$ par celle d'aires de rectangles dont la base est $1/2N$ et la hauteur est $(\Pi^t)_{[i,j]}$. Par le lemme C.0.1, cette somme peut être approximée par l'intégrale de $(Pe^{PRP\tau})_{[i,j]}$ sur $\tau \geq 0$ lorsque N tend vers l'infini en prenant $t = [2N\tau]$. Pour déterminer quels éléments des matrices P et $e^{PRP\tau}$ doivent être utilisés dans l'intégrale, on doit bien savoir ce que représente ces deux matrices. L'annexe C décrit plus en profondeur les caractéristiques de celles-ci. Ici, on se contente d'expliquer que les états 1 à 15 peuvent être séparés en deux classes. En effet, le temps passé dans les états dans lesquels il y a deux gènes ancestraux dans un même individu (états 3, 5, 7, 8, 10, 11, 13, 14 et 15) est beaucoup plus court, en moyenne, que celui passé dans les états dans lesquels les gènes ancestraux sont dans des individus différents (états 1, 2, 4, 6, 9 et 12) car, rapidement, soit les gènes ancestraux dans un même individu coalescent, soit ils se séparent dans des individus différents.

Ces deux phénomènes représentent des changements d'état qu'on dit instantanés. Le temps passé dans un état dans lequel les gènes ancestraux sont dans des individus différents est beaucoup plus long, car il faut beaucoup de temps pour que deux de ces gènes ancestraux se rencontrent dans un même individu lorsque N est grand. Ces deux classes d'états amènent deux échelles de temps différentes. La matrice P devant l'intégrale représente la première transition instantanée à partir de la configuration des gènes ancestraux du temps τ et en remontant le temps, sachant que l'allèle A n'est présent qu'une fois à la génération 0. La matrice $e^{PRP\tau}$ représente le reste du processus constitué de passages à plus longs termes (voir figure 3.4).

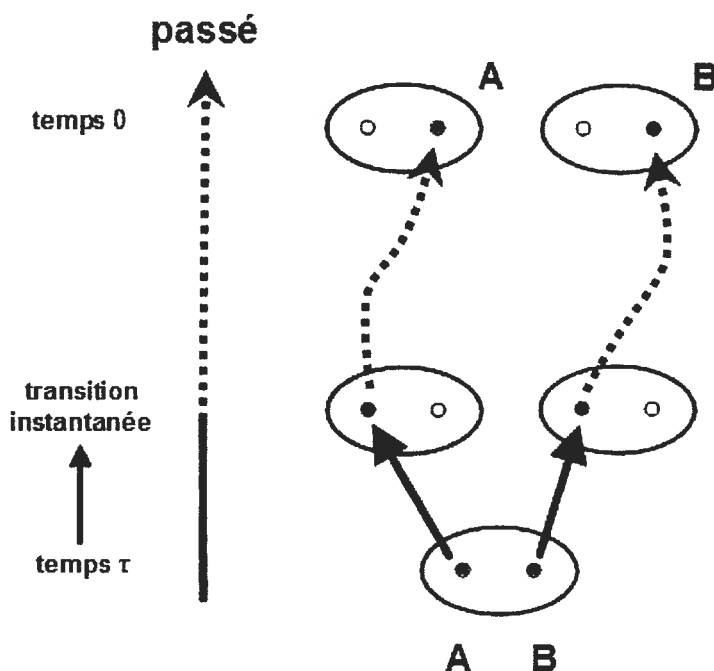


FIG. 3.4. Façon d'avoir la génotype AB au temps τ en nombre de $2N$ générations lorsque N tend vers l'infini.

Dans le cas ci-haut par exemple, on a deux gènes ancestraux (un A et un B) dans un individu au temps τ . On débute donc à l'état 3. Rapidement, en remontant le temps, ces deux gènes ancestraux, étant dans le même individu, vont donc soit coalescer, soit se séparer. Puisque les gènes sont de types différents, ils ne peuvent coalescer, donc ils se sépareront assurément dans deux individus

différents. La configuration de ces gènes passera donc rapidement et certainement de l'état 3 à l'état 2 d'où le fait que c'est l'élément $[3,2]$ de la matrice P (avec probabilité 1) qui est devant l'intégrale. Ensuite, ces deux lignées ne pouvant coalescer, il devrait y avoir encore deux lignées au temps 0 (dans des individus différents car on fait tendre N vers l'infini et donc la probabilité de se retrouver dans un même individu tend vers 0). On doit donc rester à l'état 2 jusqu'au temps 0 d'où le fait que c'est l'élément $[2,2]$ de la matrice $e^{PRP\tau}$ qui est utilisé. En revenant à l'équation (3.2.8), l'évaluation de la limite donne

$$\begin{aligned} \lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AB}(t)) &= \frac{2(1-\alpha)}{2-\alpha} \times \int_0^\infty e^{-2\tau/(2-\alpha)} d\tau, \\ &= \frac{2(1-\alpha)}{2-\alpha} \times \frac{2-\alpha}{2} = 1-\alpha. \end{aligned}$$

La même marche à suivre peut être appliquée pour calculer les autres sommes d'espérances. Les voilà donc, dans l'ordre d'apparition de l'équation (3.2.6).

a) On considère

$$\begin{aligned} \sum_{t \geq 0} E_0(Q_{AA}(t) - Q_{AA}^2(t)) &= \sum_{t \geq 0} E_0[Q_{AA}(t)(1 - Q_{AA}(t))] , \\ &= \sum_{t \geq 0} E_0[Q_{AA}(t)(Q_{AB}(t) + Q_{BB}(t))] , \\ &= \sum_{t \geq 0} E_0(Q_{AA}(t)Q_{AB}(t)) + \sum_{t \geq 0} E_0(Q_{AA}(t)Q_{BB}(t)) . \end{aligned}$$

De manière analogue au cas qui précède, on a

$$\begin{aligned}
& \sum_{t \geq 0} E_0(Q_{AA}(t)Q_{AB}(t)) \\
&= \sum_{t \geq 0} Pr_0(\text{l'individu } k_1 \text{ soit } AA \text{ et l'individu } k_2 \neq k_1 \text{ soit } AB \text{ à la génération } t), \\
&= \sum_{t \geq 0} \left[(\Pi^t)_{[8,2]} + (\Pi^t)_{[8,3]} \right] \times \frac{1}{2N}.
\end{aligned}$$

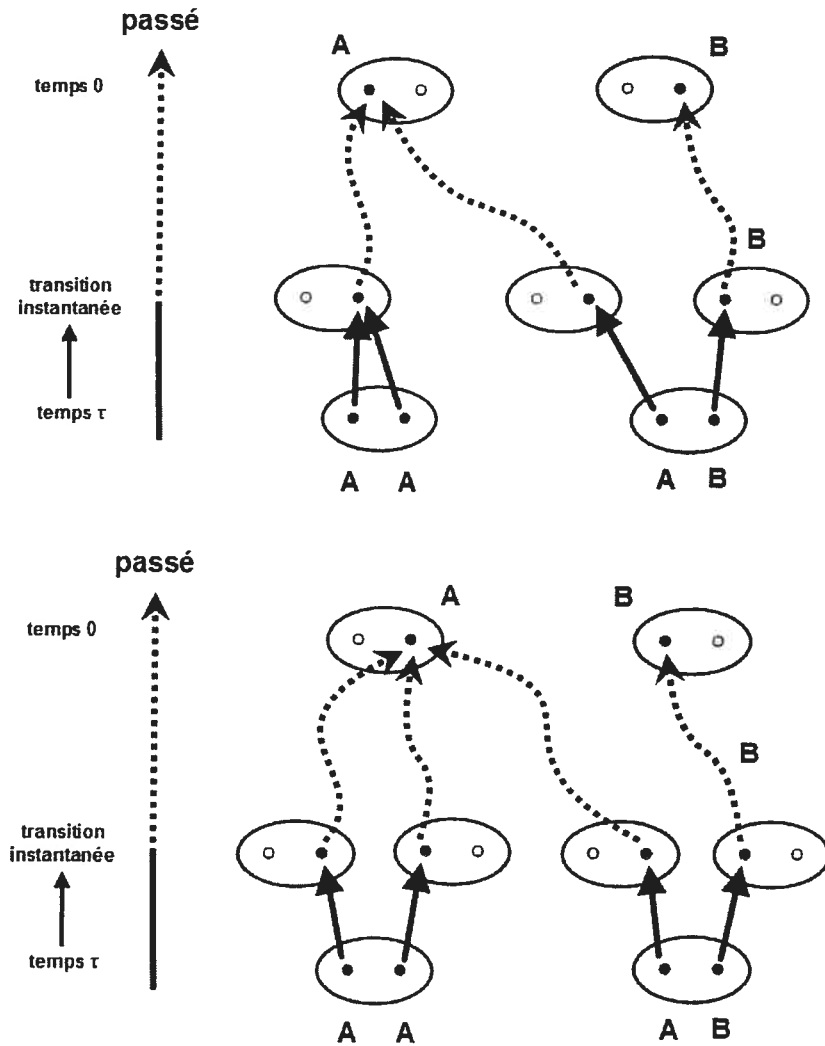


FIG. 3.5. Façons d'avoir deux individus de génotypes AA et AB au temps τ en nombre de $2N$ générations lorsque N tend vers l'infini.

Cependant, lorsque N tend vers l'infini, le terme $(\Pi^t)_{[8,3]}$ disparaît car la probabilité que deux lignées se retrouvent dans le même individu à la génération 0 tend alors vers 0. En procédant comme précédemment et en tenant compte des deux façons pour deux individus d'être de génotypes AA et AB au temps τ (figure 3.5), on obtient que

$$\begin{aligned} \lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AA}(t)Q_{AB}(t)) \\ = P_{[8,4]} \left(\mu_1 \int_0^\infty (e^{PRP\tau})_{[4,2]} d\tau \right) + P_{[8,6]} \left(\mu_2 \int_0^\infty (e^{PRP\tau})_{[6,2]} d\tau \right) . \end{aligned}$$

La configuration des gènes ancestraux est en effet à l'état 8 au temps τ car il y a quatre gènes à suivre (AA, AB) dans deux individus. Puisque l'état 8 est un état dans lequel il y a, deux fois, deux gènes ancestraux dans un même individu, il y aura rapidement une transition "instantanée" (représentée par la matrice P) à un autre état. Dans ce cas particulier, il y a deux possibilités de transitions instantanées qui sont représentées à la figure 3.5. Puisqu'il ne peut y avoir de coalescence entre A et B , les deux gènes de l'individu de génotype AB se sépareront assurément. Les deux gènes de l'individu de génotype AA peuvent soit coalescer, soit se séparer. S'ils coalescent, il y a une transition instantanée de l'état 8 à l'état 4 (ce cas est représenté par le premier terme du membre de droite de l'équation ci-haut et par le dessin du haut de la figure 3.5) alors que s'ils se séparent, la transition instantanée sera de l'état 8 à l'état 6 (ce cas est représenté par le deuxième terme du membre de droite de l'équation ci-haut et par le dessin du bas de la figure 3.5). Après la transition instantanée, les changements d'état plus long sont caractérisés par la matrice $e^{PRP\tau}$. Puisqu'il n'y a qu'un allèle A au temps 0, tous les gènes ancestraux de type A doivent coalescer ensemble, que ce soit dans l'un ou l'autre des deux cas après la transition instantanée. Puisqu'il n'y a qu'un gène ancestral de type B après la transition instantanée et que les allèles de type A ne peuvent coalescer avec les allèles de type B , la lignée ancestrale de type B doit rester jusqu'au temps 0. Il doit donc y avoir un gène ancestral de type A et un gène

ancestral de type B à la génération 0 (l'état 2) et ceci est valable dans les deux cas de transition instantanée. Cela explique le choix des éléments de la matrice $e^{PRP\tau}$.

Cependant, lorsqu'elles donnent la probabilité de passer d'une configuration à une autre par des coalescences, les matrices P et $e^{PRP\tau}$ ne tiennent pas compte des différents types de gènes (A et B) qui ne peuvent coalescer ensemble. On multiplie donc les intégrales par la probabilité de choisir deux gènes de même type pour qu'il puisse y avoir une coalescence. Ici, puisqu'il y a deux chemins possibles pour les lignées (les deux dessins de la figure 3.5), on note ces probabilités par μ_1 et μ_2 . Chacune de ces probabilités est trouvée en deux étapes, car on doit tenir compte du type des gènes ancestraux lors des deux passages (représentés par les matrices P et $e^{PRP\tau}$) dont l'échelle de temps est différente. Par exemple, pour μ_1 , lors du passage instantané ($P_{[8,4]}$), les gènes ancestraux de l'individu de génotype AB ne peuvent coalescer ensemble. Il faut donc multiplier la probabilité $P_{[8,4]}$ par la probabilité que ce soit les gènes de l'individu de type AA qui sont choisis pour coalescer et non ceux de l'individu AB . Cette probabilité est $1/2$. Ensuite, lors du passage plus long, représenté par la matrice $e^{PRP\tau}$, il faut encore que le passage de A, A, B à A, B gènes ancestraux soit compatible avec une coalescence. Il faut donc multiplier la probabilité $(e^{PRP\tau})_{[4,2]}$ par $1/3$, car il y a $1/3$ des chances de choisir le couple AA pour que la coalescence se produise. En rassemblant ces deux résultats, $\mu_1 = 1/2 \times 1/3 = 1/6$. Pour ce qui est de μ_2 , il n'y a pas de coalescence lors de la transition instantanée. Il n'y a donc pas de restriction en ce qui concerne la transition $P_{[8,6]}$ (la probabilité de la possibilité de cette transition est donc 1). Ensuite, lors du long processus, les deux coalescences doivent être entre des gènes de type A . Il y a donc $3/6$ des chances de choisir un couple de gènes de type A pour que la première coalescence se produise et ensuite, $1/3$ des chances de choisir un couple de gènes de type A pour que la deuxième coalescence se produise. Alors, $\mu_2 = 1 \times 3/6 \times 1/3 = 1/6$. L'évaluation de la limite donne

$$\begin{aligned}
& \frac{4\alpha(1-\alpha)}{[2-\alpha]^2} \left(\frac{1}{6} \times \int_0^\infty \frac{3}{2} (e^{-2\tau/(2-\alpha)}) (1 - e^{-4\tau/(2-\alpha)}) d\tau \right) \\
& + \left[\frac{2(1-\alpha)}{2-\alpha} \right]^2 \left(\frac{1}{6} \times \int_0^\infty \frac{3}{5} (e^{-2\tau/(2-\alpha)}) (3 - 5 e^{-4\tau/(2-\alpha)} + 2 e^{-10\tau/(2-\alpha)}) d\tau \right) \\
& = \frac{4\alpha(1-\alpha)}{[2-\alpha]^2} \left(\frac{1}{6} \times \frac{2-\alpha}{2} \right) + \left[\frac{2(1-\alpha)}{2-\alpha} \right]^2 \left(\frac{1}{6} \times \frac{2-\alpha}{2} \right) , \\
& = \frac{\alpha(1-\alpha)}{3(2-\alpha)} + \frac{[2(1-\alpha)]^2}{12(2-\alpha)} . \tag{3.2.9}
\end{aligned}$$

De manière analogue et en se référant aux figures 3.6 et 3.7, on a

$$\begin{aligned}
& \lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AA}(t)Q_{BB}(t)) \\
& = P_{[8,2]} \int_0^\infty (e^{PRP\tau})_{[2,2]} d\tau + 2 P_{[8,4]} \left(\mu_3 \int_0^\infty (e^{PRP\tau})_{[4,2]} d\tau \right) \\
& + P_{[8,6]} \left(\mu_4 \int_0^\infty (e^{PRP\tau})_{[6,2]} d\tau \right) + P_{[8,4]} \left(\mu_5 \int_0^\infty (e^{PRP\tau})_{[4,4]} d\tau \right) \\
& + P_{[8,6]} \left(\mu_6 \int_0^\infty (e^{PRP\tau})_{[6,4]} d\tau \right) . \tag{3.2.10}
\end{aligned}$$

En effet, ces termes peuvent s'associer aux six façons, représentées par les figures 3.6 et 3.7, d'avoir deux individus de génotypes AA et BB au temps τ sachant que la fréquence de A au temps 0 est $1/2N$. Dans tous les cas, l'état dans lequel les gènes ancestraux se trouvent au temps τ est l'état 8, car il y a deux individus avec deux gènes ancestraux. Le premier terme du membre de droite de l'équation (3.2.10) est associé avec le dessin du haut de la figure 3.6. Dans ce cas, il y a un passage instantané de l'état 8 à l'état 2 ($P_{[8,2]}$), qui est possible car les deux gènes de chaque individu sont de même type donc ils peuvent coalescer. Ensuite,

les deux gènes ancestraux restant étant de type différent, la seule possibilité pour la suite est de rester à l'état deux jusqu'au temps 0 (d'où $(e^{PRP\tau})_{[2,2]}$). Comme le type des gènes ancestraux ne donne pas de restriction sur la possibilité des coalescences, il n'est pas nécessaire de multiplier par une probabilité μ . Le deuxième terme de l'équation (3.2.10) est associé avec les deux dessins du milieu et du bas de la figure 3.6. Ces deux dessins représentent deux possibilités de chemins symétriques (il n'y a que le type des gènes ancestraux qui change). Donc, plutôt que de trouver ces deux probabilités de possibilités de chemin et de les additionner, on peut se permettre de n'en trouver qu'une pour ensuite multiplier cette probabilité par deux pour aussi représenter l'autre possibilité. Référons-nous au dessin du milieu de la figure 3.6 par exemple. D'abord, la transition instantanée se fait de l'état 8 à l'état 4 ($P_{[8,4]}$). En considérant les génotypes, la probabilité pour que ce soit les gènes de l'individu BB qui coalescent lors de la transition instantanée (probabilité qui fait partie de μ_3), et non pas ceux de l'individu AA , est égale à $1/2$. Pour le reste du chemin, il y a passage de l'état 4 à l'état 2 (d'où $(e^{PRP\tau})_{[4,2]}$) par une coalescence. En tenant compte du type des gènes, ce dernier passage n'est possible que si ce sont les deux gènes ancestraux de type A , parmi les gènes A, A, B , qui coalescent ensemble et cette probabilité est $1/3$. On obtient donc que $\mu_3 = 1/2 \times 1/3 = 1/6$.

Le troisième terme de l'équation (3.2.10) représente la possibilité de trajet des lignées ancestrales illustrée par le dessin du haut de la figure 3.7. La transition instantanée se fait de l'état 8 à l'état 6 ($P_{[8,6]}$), car les deux gènes des deux individus se séparent, ce qui ne dépend pas du type des gènes. Ensuite, alors que chaque gène ancestral est seul dans un individu, le reste du chemin se fait de l'état 6 à l'état 2 (d'où $(e^{PRP\tau})_{[6,2]}$) par deux coalescences. Ces deux coalescences ne pouvant se produire entre n'importe quels gènes, il faut tenir compte de la probabilité μ_4 qui représente la probabilité d'une coalescence entre les deux gènes de type A et une autre entre les deux gènes de type B . On trouve alors $\mu_4 = 1/9$, car il y a deux paires de gènes ancestraux sur les six paires possibles, à partir de A, A, B, B , qui peuvent coalescer ensemble et ensuite, pour la deuxième

coalescence, il reste une paire sur trois ce qui fait $2/6 \times 1/3 = 1/9$. Le quatrième terme de l'équation (3.2.10) représente le dessin du milieu de la figure 3.7. On remarque qu'il est en effet possible d'avoir plus qu'un gène ancestral de type B au temps 0 (la fréquence de B est $1 - 1/2N$). L'important, c'est qu'il n'y ait pas deux gènes ancestraux dans le même individu car cette probabilité tend vers 0 lorsque N tend vers l'infini (c'est le cas ici). Considérant ce terme, on a d'abord une transition instantanée dans laquelle il y a un passage de l'état 8 à l'état 4 ($P_{[8,4]}$). Cependant, pour avoir deux gènes ancestraux de type B au temps 0, il faut absolument que ce soit les gènes de l'individu de génotype AA du temps τ qui coalescent ensemble. La probabilité de cela est $1/2$ et fait partie de μ_5 . Ensuite, avec une échelle de temps plus grande, les gènes ancestraux restent à l'état 4 jusqu'au temps 0 (d'où $(e^{PRP\tau})_{[4,4]}$). Puisqu'il n'y a pas de coalescence lors de cette dernière étape, le type des individus n'y apporte aucune restriction et donc, $\mu_5 = 1/2$. Le dernier terme de l'équation (3.2.10) est associé avec le dessin du bas de la figure 3.7. Il y a d'abord une transition instantanée de l'état 8 à l'état 6 ($P_{[8,6]}$) pour laquelle il n'y a pas de coalescence, donc le type des gènes n'a pas à être considéré. Ensuite, dans une autre échelle de temps, il y a coalescence et passage de l'état 6 à l'état 4 (d'où $(e^{PRP\tau})_{[6,4]}$). Cette coalescence peut se produire seulement entre des gènes de même type. Puisqu'il reste encore deux gènes ancestraux de type B au temps 0, la coalescence ne peut se produire qu'entre les deux gènes de type A parmi les paires possibles avec les gènes A, A, B, B . Cela fait donc une probabilité de $1/6$ d'où $\mu_6 = 1/6$. Notons que les intégrales de $(e^{PRP\tau})_{[2,2]}$, $(e^{PRP\tau})_{[4,2]}$ et $(e^{PRP\tau})_{[6,2]}$ avaient déjà été trouvées un peu plus haut. En évaluant la limite, on a

$$\lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AA}(t)Q_{BB}(t))$$

$$\begin{aligned}
&= \left[\frac{\alpha}{2-\alpha} \right]^2 \left(\frac{2-\alpha}{2} \right) + 2 \times \frac{4\alpha(1-\alpha)}{[2-\alpha]^2} \left(\frac{1}{6} \times \frac{2-\alpha}{2} \right) \\
&\quad + \left[\frac{2(1-\alpha)}{2-\alpha} \right]^2 \left(\frac{1}{9} \times \frac{2-\alpha}{2} \right) + \frac{4\alpha(1-\alpha)}{[2-\alpha]^2} \left(\frac{1}{2} \times \int_0^\infty e^{-6\tau/(2-\alpha)} d\tau \right) \\
&\quad + \left[\frac{2(1-\alpha)}{2-\alpha} \right]^2 \left(\frac{1}{6} \times \int_0^\infty 2 (e^{-6\tau/(2-\alpha)}) (1 - e^{-6\tau/(2-\alpha)}) d\tau \right) , \\
&= \left[\frac{\alpha}{2-\alpha} \right]^2 \left(\frac{2-\alpha}{2} \right) + \frac{8\alpha(1-\alpha)}{[2-\alpha]^2} \left(\frac{1}{6} \times \frac{2-\alpha}{2} \right) + \left[\frac{2(1-\alpha)}{2-\alpha} \right]^2 \left(\frac{1}{9} \times \frac{2-\alpha}{2} \right) \\
&\quad + \frac{4\alpha(1-\alpha)}{[2-\alpha]^2} \left(\frac{1}{2} \times \frac{2-\alpha}{6} \right) + \left[\frac{2(1-\alpha)}{2-\alpha} \right]^2 \left(\frac{1}{6} \times \frac{2-\alpha}{6} \right) , \\
&= \frac{\alpha^2}{2(2-\alpha)} + \frac{\alpha(1-\alpha)}{2-\alpha} + \frac{[2(1-\alpha)]^2}{12(2-\alpha)} . \tag{3.2.11}
\end{aligned}$$

En additionnant maintenant les équations (3.2.9) et (3.2.11), on obtient :

$$\begin{aligned}
&\lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AA}(t) - Q_{AA}^2(t)) \\
&= \frac{\alpha(1-\alpha)}{3(2-\alpha)} + \frac{[2(1-\alpha)]^2}{12(2-\alpha)} + \frac{\alpha^2}{2(2-\alpha)} + \frac{\alpha(1-\alpha)}{2-\alpha} + \frac{[2(1-\alpha)]^2}{12(2-\alpha)} , \\
&= \frac{\alpha^2}{2(2-\alpha)} + \frac{4\alpha(1-\alpha)}{3(2-\alpha)} + \frac{[2(1-\alpha)]^2}{6(2-\alpha)} .
\end{aligned}$$

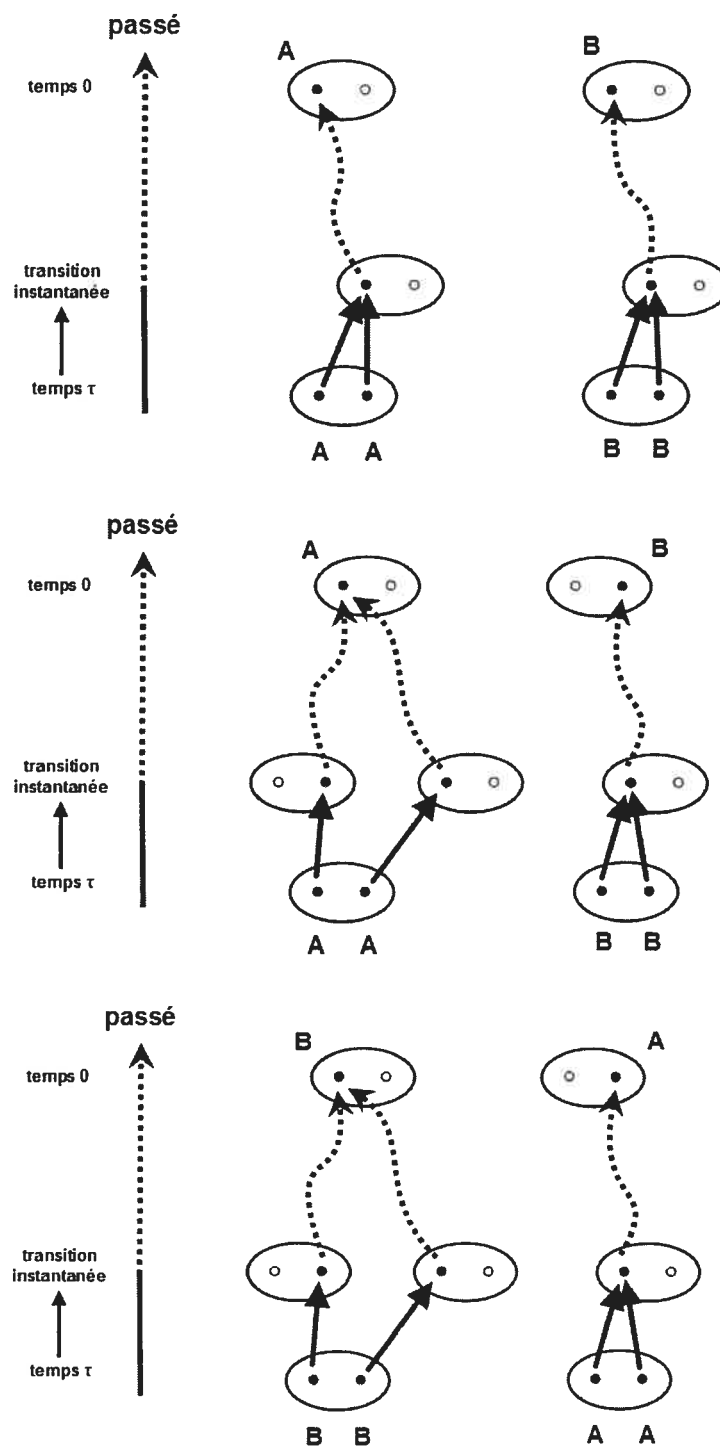


FIG. 3.6. Trois façons d'avoir deux individus de génotypes AA et BB au temps τ en nombre de $2N$ générations lorsque N tend vers l'infini.

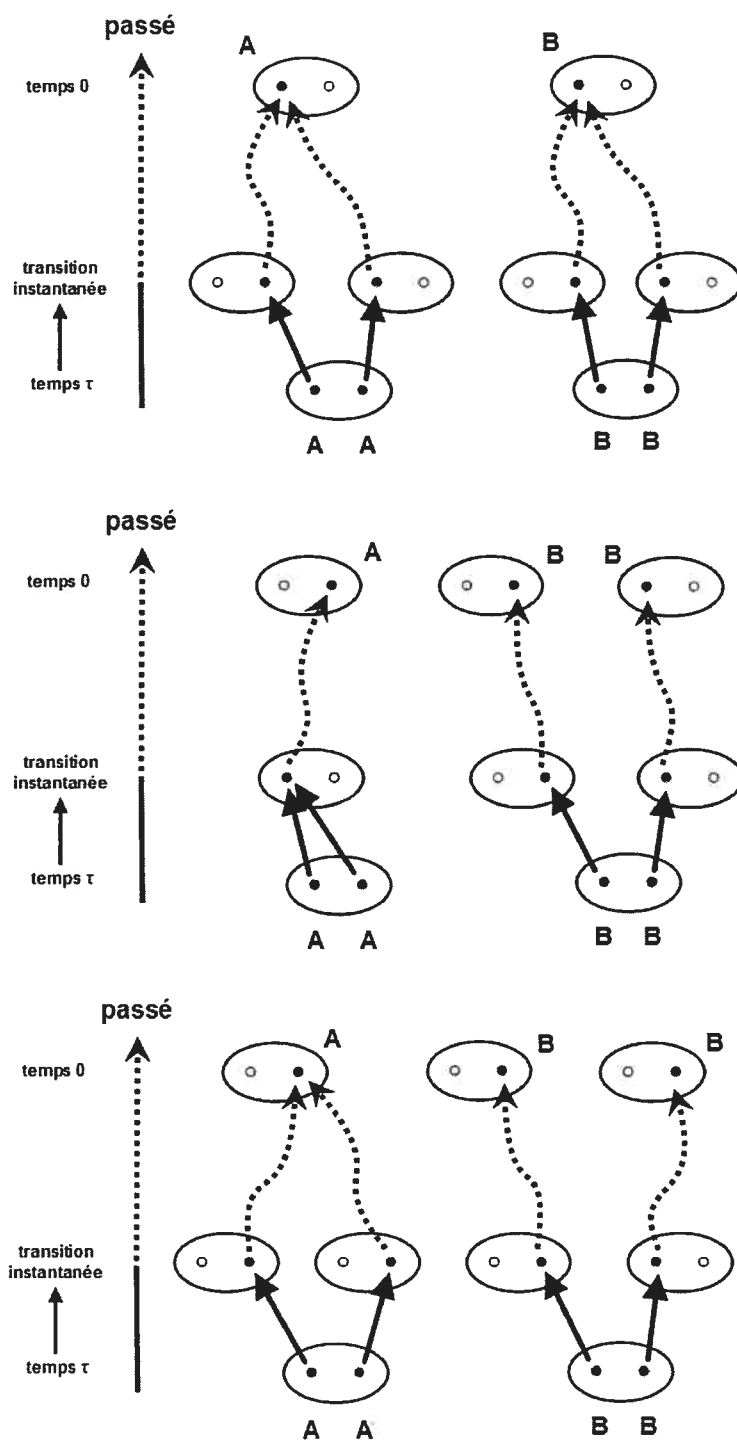


FIG. 3.7. Trois autres façons d'avoir deux individus de génotypes AA et BB au temps τ en nombre de $2N$ générations lorsque N tend vers l'infini.

b) On considère par la suite

$$\begin{aligned}
 \sum_{t \geq 0} E_0(Q_{AA}^2(t) - Q_{AA}^3(t)) &= \sum_{t \geq 0} E_0[Q_{AA}^2(t)(1 - Q_{AA}(t))] , \\
 &= \sum_{t \geq 0} E_0[Q_{AA}^2(t)(Q_{AB}(t) + Q_{BB}(t))] , \\
 &= \sum_{t \geq 0} E_0(Q_{AA}^2(t)Q_{AB}(t)) + \sum_{t \geq 0} E_0(Q_{AA}^2(t)Q_{BB}(t)) .
 \end{aligned}$$

On a d'abord, en utilisant le lemme C.0.1 de Möhle, que

$$\begin{aligned}
 \lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AA}^2(t)Q_{AB}(t)) \\
 &= P_{[15,12]} \left(\mu_7 \int_0^\infty (e^{PRP\tau})_{[12,2]} d\tau \right) + P_{[15,9]} \left(\mu_8 \int_0^\infty (e^{PRP\tau})_{[9,2]} d\tau \right) \\
 &\quad + P_{[15,6]} \left(\mu_9 \int_0^\infty (e^{PRP\tau})_{[6,2]} d\tau \right) . \tag{3.2.12}
 \end{aligned}$$

Il y a en effet trois possibilités de chemins qui constituent la probabilité qu'il y ait deux individus AA et un AB au temps τ sachant qu'il n'y a qu'un gène de type A au temps 0. Dans tous les cas, l'état des gènes ancestraux au temps τ est 15 car on doit suivre les lignées de six gènes de trois individus (voir la figure 3.2). Le premier terme du membre de droite de l'équation (3.2.12) correspond aux possibilités de lignées dans lesquelles il y a une transition instantanée de l'état 15 à l'état 12 ($P_{[15,12]}$) ce qui implique qu'il n'y a pas de coalescence instantanée et que les trois paires de gènes ancestraux se séparent dans six individus différents. Puisque le type des gènes ancestraux ne change pas la possibilité de réaliser cette transition, on n'a pas à multiplier par une probabilité qui serait incluse dans μ_7 . Suite à la transition instantanée, il reste six gènes ancestraux A, A, A, A, A, B . Ensuite, la seule façon d'arriver au temps 0 avec seulement un gène ancestral de type A est que tous les gènes de type A coalescent entre eux. C'est un passage de l'état 12 à l'état 2 (d'où $(e^{PRP\tau})_{[12,2]}$). Il faut cependant multiplier par la

probabilité de choisir, à quatre reprises, une paire de A , parmi les autres paires de gènes ancestraux, pour qu'il y ait coalescence. On trouve que cette probabilité est égale à $10/15 \times 6/10 \times 3/6 \times 1/3 = 1/15$ et donc, $\mu_7 = 1/15$.

Le deuxième terme du membre de droite de l'équation (3.2.12) correspond aux possibilités de lignées dans lesquelles les deux gènes de type A d'un des individus AA du temps τ coalescent rapidement alors que les quatre autres gènes se séparent. C'est un passage instantané de l'état 15 à l'état 9 ($P_{[15,9]}$). Puisqu'il y a deux individus de génotype AA au temps τ et que l'un ou l'autre peut être celui dont les gènes coalescent mais qu'on ne peut choisir les gènes de l'individu de génotype AB pour coalescer, on doit donc multiplier $P_{[15,9]}$ par $2/3$, où $2/3$ fait partie de μ_8 . Par la suite, l'échelle de temps devient plus grande et le passage jusqu'au temps 0 se fait de l'état 9 à l'état 2 (d'où $(e^{PRP\tau})_{[9,2]}$) par trois coalescences. Il faut donc multiplier par la probabilité de choisir, à trois reprises, une paire de A , parmi les autres paires de gènes ancestraux, pour qu'il y ait trois coalescences de gènes de type A . On trouve que cette probabilité est égale à $6/10 \times 3/6 \times 1/3 = 1/10$ et alors, on conclut que $\mu_8 = 2/3 \times 1/10 = 1/15$. Enfin, le dernier terme à droite de l'équation (3.2.12) est associé aux possibilités de lignées dans lesquelles il y a une transition instantanée de l'état 15 à l'état 6 ($P_{[15,6]}$) qui amène, toujours en remontant le temps, aux gènes ancestraux A, A, A, B . Puisqu'il y a 1 chance sur 3 que ce soit les deux individus de génotype AA qui soient choisis pour coalescer, ce $1/3$ contribue donc à μ_9 . Ensuite, par un processus plus long et jusqu'au temps 0, il ne peut y avoir qu'un passage de l'état 6 à l'état 2 (d'où $(e^{PRP\tau})_{[6,2]}$) qui ne peut être que la cause de deux coalescences. Pour que cela soit possible, il faut que ces coalescences soient parmi les gènes de type A . On multiplie donc par la probabilité de choisir une paire de A , à deux reprises, parmi les autres paires de gènes ancestraux restant. On trouve que cette probabilité est égale à $3/6 \times 1/3 = 1/6$, et donc, que $\mu_9 = 1/3 \times 1/6 = 1/18$. On peut maintenant poursuivre le calcul et puisque l'intégrale de $(e^{PRP\tau})_{[6,2]}$ avait déjà été trouvée dans le calcul de la limite lorsque N tend vers l'infini de $\sum_{t \geq 0} E_0(Q_{AA}(t)Q_{AB}(t))$, l'évaluation de la limite donne

$$\left[\frac{2(1-\alpha)}{2-\alpha} \right]^3 \left(\frac{1}{15} \times I_1 \right) + \frac{12(1-\alpha)^2 \alpha}{(2-\alpha)^3} \left(\frac{1}{15} \times I_2 \right) + \frac{6(1-\alpha) \alpha^2}{(2-\alpha)^3} \left(\frac{1}{18} \times \frac{2-\alpha}{2} \right) ,$$

où

$$I_1 = \int_0^\infty \frac{5}{14} \left(6e^{-2\tau/(2-\alpha)} - 15e^{-6\tau/(2-\alpha)} + 14e^{-12\tau/(2-\alpha)} - 6e^{-20\tau/(2-\alpha)} + e^{-30\tau/(2-\alpha)} \right) d\tau ,$$

et

$$I_2 = \int_0^\infty \left(2e^{-2\tau/(2-\alpha)} - \frac{30}{7}e^{-6\tau/(2-\alpha)} + 3e^{-12\tau/(2-\alpha)} - \frac{5}{7}e^{-20\tau/(2-\alpha)} \right) d\tau .$$

On a donc que

$$\begin{aligned} \lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AA}^2(t) Q_{AB}(t)) \\ = \left[\frac{2(1-\alpha)}{2-\alpha} \right]^3 \left(\frac{1}{15} \times \frac{2-\alpha}{2} \right) + \frac{12(1-\alpha)^2 \alpha}{(2-\alpha)^3} \left(\frac{1}{15} \times \frac{2-\alpha}{2} \right) \\ + \frac{6(1-\alpha) \alpha^2}{(2-\alpha)^3} \left(\frac{1}{18} \times \frac{2-\alpha}{2} \right) , \\ = \frac{[2(1-\alpha)]^3}{30(2-\alpha)^2} + \frac{\alpha[2(1-\alpha)]^2}{10(2-\alpha)^2} + \frac{\alpha^2(1-\alpha)}{6(2-\alpha)^2} . \end{aligned} \quad (3.2.13)$$

Avec la même méthode, on a

$$\begin{aligned}
& \lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AA}^2(t)Q_{BB}(t)) \\
&= P_{[15,12]} \left(\mu_{10} \int_0^\infty (e^{PRP\tau})_{[12,4]} d\tau \right) + P_{[15,9]} \left(\mu_{11} \int_0^\infty (e^{PRP\tau})_{[9,4]} d\tau \right) \\
&\quad + P_{[15,6]} \left(\mu_{12} \int_0^\infty (e^{PRP\tau})_{[6,4]} d\tau \right) + P_{[15,12]} \left(\mu_{13} \int_0^\infty (e^{PRP\tau})_{[12,2]} d\tau \right) \\
&\quad + P_{[15,9]} \left(\mu_{14} \int_0^\infty (e^{PRP\tau})_{[9,2]} d\tau \right) + P_{[15,6]} \left(\mu_{15} \int_0^\infty (e^{PRP\tau})_{[6,2]} d\tau \right) \\
&\quad + P_{[15,9]} \left(\mu_{16} \int_0^\infty (e^{PRP\tau})_{[9,2]} d\tau \right) + P_{[15,6]} \left(\mu_{17} \int_0^\infty (e^{PRP\tau})_{[6,2]} d\tau \right) \\
&\quad + P_{[15,4]} \left(\mu_{18} \int_0^\infty (e^{PRP\tau})_{[4,2]} d\tau \right) . \tag{3.2.14}
\end{aligned}$$

Les termes du membre de droite de cette équation représentent toutes les possibilités de chemins pour des lignées ancestrales, de deux individus de type AA et un de type BB au temps τ , sachant qu'il n'y a qu'un allèle A au temps 0. Dans tous les cas, l'état des gènes ancestraux à la génération τ est 15 car on doit suivre les six gènes appartenant à trois individus. Le premier terme de l'équation (3.2.14) correspond aux possibilités de lignées dans lesquelles il y a une transition instantanée de l'état 15 à l'état 12 ($P_{[15,12]}$) et ensuite un lent passage de l'état 12 vers l'état 4 au temps 0. Il n'y a donc aucune coalescence instantanée et les trois paires de gènes ancestraux se séparent dans six individus différents. Puisque la possibilité de réaliser cette transition ne change pas avec le type des gènes ancestraux, on ne multiplie pas par une probabilité qui serait incluse dans μ_{10} . Suite à la transition instantanée, il reste six gènes ancestraux A, A, A, A, B, B dans six individus différents. Ensuite, la seule façon d'arriver au temps 0 avec seulement un gène ancestral de type A et deux de type B est que tous les gènes de type A coalescent entre eux et que ceux de type B ne coalescent pas. C'est un passage de l'état 12 à l'état 4 (d'où $(e^{PRP\tau})_{[12,4]}$). Il faut cependant multiplier par la probabilité de choisir, à trois reprises, une paire de A , parmi les autres paires de gènes ancestraux, pour qu'il y ait des coalescences compatibles. On trouve que

cette probabilité est égale à $6/15 \times 3/10 \times 1/6 = 1/50$ et donc, $\mu_{10} = 1/50$.

Le deuxième terme du membre de droite de l'équation (3.2.14) correspond aux possibilités de lignées dans lesquelles les deux gènes de type A d'un des individus AA du temps τ coalescent rapidement alors que les quatre autres gènes se séparent, pour aboutir à A, B, B au temps 0. C'est une transition instantanée de l'état 15 à l'état 9 ($P_{[15,9]}$). Puisqu'il y a deux individus de génotype AA au temps τ et que l'un ou l'autre peut être celui dont les gènes coalescent mais qu'il ne faut pas que ce soit les gènes de l'individu de génotype BB qui soient choisis pour coalescer, on doit multiplier $P_{[15,9]}$ par $2/3$. Cette dernière probabilité fait partie de μ_{11} . Par la suite, l'échelle de temps devient plus grande et le passage jusqu'au temps 0 se fait de l'état 9, avec A, A, A, B, B , à l'état 4, avec A, B, B , (d'où $(e^{PRP\tau})_{[9,4]}$) par deux coalescences entre des gènes ancestraux de type A . Il faut donc multiplier par la probabilité de choisir, à deux reprises, une paire de A , parmi les autres paires de gènes ancestraux, pour qu'il y ait deux coalescences de gènes de type A . On trouve que cette probabilité est égale à $3/10 \times 1/6 = 1/20$. On obtient donc que $\mu_{11} = 2/3 \times 1/20 = 1/30$. Le troisième terme de droite de l'équation (3.2.14) est associé aux possibilités de lignées dans lesquelles il y a une transition instantanée de l'état 15 à l'état 6 ($P_{[15,6]}$) qui amène aux gènes ancestraux A, A, B, B , et ensuite, un passage jusqu'au temps 0 avec A, B, B . Puisqu'il y a 1 chance sur 3 que ce soit les deux individus de génotype AA qui soient choisis pour coalescer instantanément, ce $1/3$ contribue donc à μ_{12} . Ensuite, par le processus plus long et jusqu'au temps 0, il y a un passage de l'état 6 à l'état 4 (d'où $(e^{PRP\tau})_{[6,4]}$) qui ne peut être que la cause d'une seule coalescence. Pour que cela soit possible, il faut que cette coalescence soit parmi les gènes de type A . On multiplie donc par la probabilité de choisir une paire de gènes de type A parmi les paires possibles de l'ensemble A, A, B, B . On trouve que cette probabilité est égale à $1/6$, et donc, que $\mu_{12} = 1/3 \times 1/6 = 1/18$.

Le quatrième terme de l'équation (3.2.14) correspond aux possibilités de lignées dans lesquelles il y a une transition instantanée de l'état 15 à l'état 12 ($P_{[15,12]}$)

et ensuite un lent passage de l'état 12 vers l'état 2 au temps 0. Donc, lors de la transition instantanée, les trois paires de gènes ancestraux se séparent dans six individus différents. La possibilité de réaliser cette transition ne change pas avec le type des gènes ancestraux, alors μ_{13} ne sera pas affecté par la transition instantanée. Suite à celle-ci, il reste six gènes ancestraux A, A, A, A, B, B . Ensuite, lors du processus plus long, la seule façon d'arriver au temps 0 avec seulement un gène ancestral de type A et un de type B , est que tous les gènes de type A coalescent entre eux et que les deux gènes de type B le fassent aussi. C'est un passage de l'état 12 à l'état 2 (d'où $(e^{PRP\tau})_{[12,2]}$). Cependant, il faut multiplier par la probabilité que le passage de A, A, A, A, B, B à A, B gènes ancestraux soit compatible avec quatre coalescences. Puisque cette probabilité est $2/75$, on a que $\mu_{13} = 2/75$. Le cinquième terme à droite de l'équation (3.2.14) correspond aux possibilités de lignées dans lesquelles il y a un passage instantané de AA, AA, BB à A, A, A, B, B ($P_{[15,9]}$), et un passage plus lent de A, A, A, B, B à A, B (d'où $(e^{PRP\tau})_{[9,2]}$). On doit multiplier $P_{[15,9]}$ par $2/3$ car il y a deux chances sur trois que ce soit un individu de génotype AA , qui soit choisis parmi les individus du temps τ , dont les gènes coalescent instantanément. Cette probabilité fait partie de μ_{14} . Le dernier passage jusqu'au temps 0 se fait de l'état 9 à l'état 2 par trois coalescences entre des gènes ancestraux de même type. Il faut donc multiplier par la probabilité de choisir, à trois reprises, une paire de gènes ancestraux de même type parmi toutes les paires possibles, pour qu'il y ait deux coalescences de gènes de type A et une de gènes de type B . On trouve que cette probabilité est égale à $1/20$. On obtient donc que $\mu_{14} = 2/3 \times 1/20 = 1/30$.

Le sixième terme de l'équation (3.2.14) est associé aux possibilités de lignées dans lesquelles il y a une transition instantanée de l'état 15 à l'état 6 ($P_{[15,6]}$) qui amène aux gènes ancestraux A, A, B, B , suivie d'une lente transition vers A, B au temps 0. Puisqu'il y a 1 chance sur 3 que ce soit les deux individus de génotype AA , au temps τ , qui soient choisis pour coalescer instantanément, ce $1/3$ contribue donc à μ_{15} . Ensuite, par le processus plus long et jusqu'au temps 0, il y a un passage de

l'état 6 à l'état 2 (d'où $(e^{PRP\tau})_{[6,2]}$) qui ne peut être que la cause de deux coalescences. Pour que cela soit possible, il faut que ces coalescences soient parmi des gènes de même type. On multiplie donc par la probabilité de choisir des gènes de même type, parmi les paires possibles avant chaque coalescence, pour que le passage de A, A, B, B à A, B gènes ancestraux soit compatible avec deux coalescences. On trouve que cette probabilité est $1/9$, et donc, que $\mu_{15} = 1/3 \times 1/9 = 1/27$. Le septième terme de l'équation (3.2.14) correspond aux possibilités de lignées dans lesquelles il y a une transition instantanée de l'état 15, avec AA, AA, BB , à l'état 9, avec A, A, A, A, B , ($P_{[15,9]}$), par une coalescence des gènes de l'individu de génotype BB au temps τ , et ensuite un lent passage de l'état 9 vers l'état 2, avec A, B , au temps 0. Puisque la probabilité de choisir l'individu de génotype BB parmi les trois au temps τ est $1/3$, la part de la transition instantanée dans μ_{16} est $1/3$. Suite à la transition instantanée, il reste les gènes ancestraux A, A, A, A, B . La seule façon d'arriver au temps 0 avec seulement un gène ancestral de type A , et un de type B , est que tous les gènes de type A coalescent entre eux. C'est donc un passage de l'état 9 à l'état 2 (d'où $(e^{PRP\tau})_{[9,2]}$). Il faut cependant multiplier par la probabilité de choisir, à trois reprises, une paire de A , parmi les autres paires de gènes ancestraux, pour que les coalescences puissent se produire. On trouve que cette probabilité est égale à $1/10$ et donc, $\mu_{16} = 1/3 \times 1/10 = 1/30$.

Le huitième terme de l'équation (3.2.14) est associé aux possibilités de lignées dans lesquelles il y a une transition instantanée de l'état 15 à l'état 6 ($P_{[15,6]}$) qui amène aux gènes ancestraux A, A, A, B et ensuite un lent passage vers A, B au temps 0. Puisqu'il y a 2 chances sur 3 que les deux individus, au temps τ , qui soient choisis pour coalescer instantanément soient de génotype AA et BB , ce $2/3$ contribue donc à μ_{17} . Par la suite, lors du processus plus long et jusqu'au temps 0, il y a un passage de l'état 6 à l'état 2 (d'où $(e^{PRP\tau})_{[6,2]}$) qui ne peut être que la cause de deux coalescences. Pour que cela soit possible, il faut que ces coalescences soient parmi des gènes de type A . On multiplie donc par la probabilité de choisir des paires de gènes de type A , avant chaque coalescence, pour que le passage de A, A, A, B à A, B gènes ancestraux soit compatible avec deux coalescences.

On trouve que cette probabilité est $1/6$, et donc, que $\mu_{17} = 2/3 \times 1/6 = 1/9$. Enfin, le dernier terme du membre de droite de l'équation (3.2.14) représente les possibilités de lignées dans lesquelles il y a une transition instantanée de l'état 15 à l'état 4 ($P_{15,4}$) par trois coalescences instantanées. Le génotype des individus au temps τ n'ayant pas d'effet sur la possibilité de réaliser celles-ci, on ne multiplie pas par une probabilité qui ferait partie de μ_{18} . Après ces coalescences, les gènes ancestraux restant sont A, A, B . Ensuite, le passage plus lent jusqu'au temps 0 ne peut être que de l'état 4 à l'état 2 (d'où $(e^{PRP\tau})_{[4,2]}$) par la coalescence des deux gènes ancestraux de type A restant. La probabilité de choisir la paire de type A parmi A, A, B étant $1/3$, $\mu_{18} = 1/3$. Puisqu'on a déjà trouvé la valeur de toutes ces intégrales sauf les deux premières, l'évaluation de la limite donne

$$\begin{aligned} & \left[\frac{2(1-\alpha)}{2-\alpha} \right]^3 \left(\frac{1}{50} \times I_3 \right) + \frac{12(1-\alpha)^2\alpha}{(2-\alpha)^3} \left(\frac{1}{30} \times I_4 \right) \\ & + \frac{6(1-\alpha)\alpha^2}{(2-\alpha)^3} \left(\frac{1}{18} \times \frac{2-\alpha}{6} \right) + \left[\frac{2(1-\alpha)}{2-\alpha} \right]^3 \left(\frac{2}{75} \times \frac{2-\alpha}{2} \right) \\ & + \frac{12(1-\alpha)^2\alpha}{(2-\alpha)^3} \left(\frac{1}{30} \times \frac{2-\alpha}{2} \right) + \frac{6(1-\alpha)\alpha^2}{(2-\alpha)^3} \left(\frac{1}{27} \times \frac{2-\alpha}{2} \right) \\ & + \frac{12(1-\alpha)^2\alpha}{(2-\alpha)^3} \left(\frac{1}{30} \times \frac{2-\alpha}{2} \right) + \frac{6(1-\alpha)\alpha^2}{(2-\alpha)^3} \left(\frac{1}{9} \times \frac{2-\alpha}{2} \right) \\ & + \left[\frac{\alpha}{2-\alpha} \right]^3 \left(\frac{1}{3} \times \frac{2-\alpha}{2} \right) , \end{aligned}$$

où

$$I_3 = \int_0^\infty \frac{5 e^{-6\tau/(2-\alpha)}}{21} \left(-35 e^{-6\tau/(2-\alpha)} + 27 e^{-14\tau/(2-\alpha)} - 7 e^{-24\tau/(2-\alpha)} + 15 \right) d\tau ,$$

et

$$I_4 = \int_0^\infty \frac{5}{7} (e^{-6\tau/(2-\alpha)}) \left(-7 e^{-6\tau/(2-\alpha)} + 3 e^{-14\tau/(2-\alpha)} + 4 \right) d\tau .$$

On obtient donc que

$$\begin{aligned}
& \lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AA}^2(t)Q_{BB}(t)) \\
&= \left[\frac{2(1-\alpha)}{2-\alpha} \right]^3 \left(\frac{1}{50} \times \frac{2-\alpha}{6} \right) + \frac{12(1-\alpha)^2\alpha}{(2-\alpha)^3} \left(\frac{1}{30} \times \frac{2-\alpha}{6} \right) \\
&\quad + \frac{6(1-\alpha)\alpha^2}{(2-\alpha)^3} \left(\frac{1}{18} \times \frac{2-\alpha}{6} \right) + \left[\frac{2(1-\alpha)}{2-\alpha} \right]^3 \left(\frac{2}{75} \times \frac{2-\alpha}{2} \right) \\
&\quad + \frac{12(1-\alpha)^2\alpha}{(2-\alpha)^3} \left(\frac{1}{30} \times \frac{2-\alpha}{2} \right) + \frac{6(1-\alpha)\alpha^2}{(2-\alpha)^3} \left(\frac{1}{27} \times \frac{2-\alpha}{2} \right) \\
&\quad + \frac{12(1-\alpha)^2\alpha}{(2-\alpha)^3} \left(\frac{1}{30} \times \frac{2-\alpha}{2} \right) + \frac{6(1-\alpha)\alpha^2}{(2-\alpha)^3} \left(\frac{1}{9} \times \frac{2-\alpha}{2} \right) \\
&\quad + \left[\frac{\alpha}{2-\alpha} \right]^3 \left(\frac{1}{3} \times \frac{2-\alpha}{2} \right) , \\
&= \frac{[2(1-\alpha)]^3}{60(2-\alpha)^2} + \frac{7\alpha}{60} \left[\frac{2(1-\alpha)}{2-\alpha} \right]^2 + \frac{(1-\alpha)}{2} \left[\frac{\alpha}{2-\alpha} \right]^2 + \frac{\alpha^3}{6(2-\alpha)^2} . \quad (3.2.15)
\end{aligned}$$

En additionnant maintenant les équations (3.2.13) et (3.2.15), on obtient

$$\begin{aligned}
& \lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AA}^2(t) - Q_{AA}^3(t)) \\
&= \frac{[2(1-\alpha)]^3}{30(2-\alpha)^2} + \frac{\alpha[2(1-\alpha)]^2}{10(2-\alpha)^2} + \frac{\alpha^2(1-\alpha)}{6(2-\alpha)^2} \\
&\quad + \frac{[2(1-\alpha)]^3}{60(2-\alpha)^2} + \frac{7\alpha}{60} \left[\frac{2(1-\alpha)}{2-\alpha} \right]^2 + \frac{(1-\alpha)}{2} \left[\frac{\alpha}{2-\alpha} \right]^2 + \frac{\alpha^3}{6(2-\alpha)^2} , \\
&= \frac{[2(1-\alpha)]^3}{20(2-\alpha)^2} + \frac{13\alpha}{60} \left[\frac{2(1-\alpha)}{2-\alpha} \right]^2 + \frac{2(1-\alpha)}{3} \left[\frac{\alpha}{2-\alpha} \right]^2 + \frac{\alpha^3}{6(2-\alpha)^2} .
\end{aligned}$$

c) On connaît déjà l'approximation de la somme sur $t \geq 0$ de l'espérance de $Q_{AA}(t)Q_{AB}(t)$. En effet, lors de l'approximation de $\sum_{t \geq 0} E_0(Q_{AA}(t) - Q_{AA}^2(t))$ plus haut, en (a), on a trouvé que

$$\lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AA}(t)Q_{AB}(t)) = \frac{\alpha(1-\alpha)}{3(2-\alpha)} + \frac{[2(1-\alpha)]^2}{12(2-\alpha)}.$$

d) Encore de manière analogue aux cas précédents, on a

$$\lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AB}^2(t)) = P_{[8,6]} \left(\mu_{19} \int_0^\infty (e^{PRP\tau})_{[6,2]} d\tau + \mu_{20} \int_0^\infty (e^{PRP\tau})_{[6,4]} d\tau \right). \quad (3.2.16)$$

Il y a ici deux possibilités de chemins qui constituent la probabilité qu'il y ait deux individus de génotype AB au temps τ sachant qu'il n'y a qu'un allèle A au temps 0. Dans ces deux cas, puisqu'on doit remonter les lignées ancestrales des deux gènes de deux individus au temps τ , la configuration de départ de ces gènes est représenté par l'état 8. La seule possibilité de transition instantanée est celle de l'état 8 à l'état 6 ($P_{[8,6]}$) car les deux gènes de chaque individu étant de type différent au temps t , il est impossible qu'il y ait coalescence. Suite à ce passage instantané, les quatre gènes ancestraux A, A, B, B sont dans quatre individus différents et ce, dans les deux cas.

Dans le premier cas (représenté par le premier terme dans la parenthèse du membre de droite de l'équation (3.2.16)), il y a deux coalescences des gènes de même type lors du processus plus lent. C'est un passage de l'état 6, avec A, A, B, B , à l'état 2, avec A, B , au temps 0 (d'où $(e^{PRP\tau})_{[6,2]}$). Puisque le type des gènes ancestraux importe lors de ces coalescences, il faut multiplier par la probabilité que le type de ceux choisis pour coalescer soit le même. Cette probabilité est $\mu_{19} = 1/9$. Le deuxième cas (représenté par le deuxième terme dans la

parenthèse du membre de droite de l'équation (3.2.16)), est la possibilité qu'il n'y ait qu'une coalescence lors du processus lent. Il y a alors un passage de l'état 6, avec A, A, B, B , à l'état 4, avec A, B, B , (d'où $(e^{PRP\tau})_{[6,4]}$), car il doit y avoir un seul gène de type A au temps 0. Il faut donc multiplier par $\mu_{20} = 1/6$ qui est la probabilité de choisir le couple de gènes de type A , parmi tous les autres couples possibles, pour que la coalescence ait lieu. Puisqu'on a déjà trouvé la valeur des deux intégrales, l'évaluation de la limite donne

$$\left[\frac{2(1-\alpha)}{2-\alpha} \right]^2 \left(\left(\frac{1}{9} \times \frac{2-\alpha}{2} \right) + \left(\frac{1}{6} \times \frac{2-\alpha}{6} \right) \right) = \frac{[2(1-\alpha)]^2}{12(2-\alpha)} .$$

e) On a déjà trouvé l'approximation de la somme sur $t \geq 0$ de l'espérance de $Q_{AA}^2(t)Q_{AB}(t)$. En effet, lors de l'approximation de $\sum_{t \geq 0} E_0(Q_{AA}^2(t) - Q_{AA}^3(t))$, en (b), on a trouvé que

$$\lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AA}^2(t)Q_{AB}(t)) = \frac{[2(1-\alpha)]^3}{30(2-\alpha)^2} + \frac{\alpha[2(1-\alpha)]^2}{10(2-\alpha)^2} + \frac{\alpha^2(1-\alpha)}{6(2-\alpha)^2} .$$

f) On approxime maintenant la septième somme de l'équation (3.2.6). On a

$$\begin{aligned} \lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AA}(t)Q_{AB}^2(t)) \\ = P_{[15,12]} \left(\mu_{21} \int_0^\infty (e^{PRP\tau})_{[12,4]} d\tau \right) + P_{[15,9]} \left(\mu_{22} \int_0^\infty (e^{PRP\tau})_{[9,4]} d\tau \right) \\ + P_{[15,12]} \left(\mu_{23} \int_0^\infty (e^{PRP\tau})_{[12,2]} d\tau \right) + P_{[15,9]} \left(\mu_{24} \int_0^\infty (e^{PRP\tau})_{[9,2]} d\tau \right) . \end{aligned} \quad (3.2.17)$$

Le premier terme du membre de droite de cette équation correspond aux possibilités de lignées dans lesquelles il y a une transition instantanée de l'état 15,

avec AA, AB, AB , à l'état 12, avec A, A, A, A, B, B ($P_{[15,12]}$) et ensuite un lent passage de cet état 12 vers l'état 4, avec A, B, B , au temps 0. Les trois paires de gènes ancestraux se séparent donc dans six individus différents lors de la transition instantanée. Ne dépendant pas du type des gènes ancestraux, la probabilité de la réalisation de cette transition est 1. Lors du processus plus long, la seule façon d'arriver au temps 0 avec seulement un gène ancestral de type A et deux de type B , est que tous les gènes de type A coalescent entre eux sans que ceux de type B le fassent. C'est un passage de l'état 12 à l'état 4 (d'où $(e^{PRP\tau})_{[12,4]}$). Il faut cependant multiplier par la probabilité que le passage de A, A, A, A, B, B à A, B, B gènes ancestraux soit compatible avec trois coalescences. Puisque cette probabilité est $1/50$, on a que $\mu_{21} = 1 \times 1/50 = 1/50$. Le deuxième terme de l'équation (3.2.17) représente les possibilités de lignées dans lesquelles les gènes de l'individu AA du temps τ coalescent rapidement alors que les quatre autres gènes se séparent. C'est une transition instantanée de l'état 15, avec les génotypes AA, AB, AB , à l'état 9, avec A, A, A, B, B ($P_{[15,9]}$). On doit alors multiplier la probabilité de cette transition par $1/3$, car il y a une chance sur trois que ce soit l'individu de génotype AA , qui soit choisi parmi les individus du temps τ , dont les gènes coalescent instantanément. Cette probabilité fait partie de μ_{22} . L'échelle de temps devient, par la suite, plus grande et le passage jusqu'au temps 0 se fait de l'état 9 à l'état 4 (d'où $(e^{PRP\tau})_{[9,4]}$) par deux coalescences entre les gènes ancestraux de type A pour aboutir avec les gènes ancestraux de type A, B, B . Il faut donc multiplier par la probabilité de choisir, avant chacune des deux coalescences, une paire de gènes ancestraux de type A , parmi toutes les paires possibles, pour que les deux coalescences soient possibles. On trouve que cette probabilité est égale à $1/20$. On obtient donc que $\mu_{22} = 1/3 \times 1/20 = 1/60$.

Le troisième terme du membre de droite de l'équation (3.2.17) est associé aux possibilités de lignées dans lesquelles il y a une transition instantanée de l'état 15, avec AA, AB, AB , à l'état 12, avec A, A, A, A, B, B , ($P_{[15,12]}$) et ensuite un lent passage de cet état 12 vers l'état 2, avec les gènes de type A, B , au temps 0. Lors de la transition instantanée, il n'y a pas de coalescence donc la probabilité

de la réalisation de cette transition est 1. Par la suite, le passage plus lent de l'état 12 vers l'état 2 (d'où $(e^{PRP\tau})_{[12,2]}$) se fait par quatre coalescences. Puisque celles-ci doivent se produire entre des gènes ancestraux de même type, on multiplie donc par la probabilité de choisir une paire de gènes de même type parmi toutes les autres paires possibles, et ce, avant chaque coalescence. On trouve que cette probabilité est $2/75$, et alors $\mu_{23} = 1 \times 2/75 = 2/75$. Enfin, le dernier terme de l'équation (3.2.17) correspond aux possibilités de lignées dans lesquelles il y a une transition instantanée de l'état 15 à l'état 9 ($P_{[15,9]}$) suivi d'un lent passage de l'état 9 vers l'état 2 au temps 0. La probabilité $P_{[15,9]}$, de AA, AB, AB à A, A, A, B, B , doit être multipliée par $1/3$, qui est la probabilité de choisir l'individu de génotype AA , parmi les trois individus au temps τ , pour qu'il y ait vraiment une coalescence possible. Lors du processus lent, le passage de l'état 9 vers l'état 2 (d'où $(e^{PRP\tau})_{[9,2]}$), de A, A, A, B, B à A, B , est caractérisé par trois coalescences entre gènes de même type. Il faut donc multiplier par $1/20$ qui est la probabilité que le passage de A, A, A, B, B à A, B soit compatible avec les trois coalescences. On a alors que $\mu_{24} = 1/3 \times 1/20 = 1/60$.

Précédemment, on a déjà trouvé la valeur des intégrales de l'équation (3.2.17). On a alors que

$$\begin{aligned}
 \lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AA}(t)Q_{AB}^2(t)) \\
 &= \left[\frac{2(1-\alpha)}{2-\alpha} \right]^3 \left(\frac{1}{50} \times \frac{2-\alpha}{6} \right) + \frac{12(1-\alpha)^2\alpha}{(2-\alpha)^3} \left(\frac{1}{60} \times \frac{2-\alpha}{6} \right) \\
 &\quad + \left[\frac{2(1-\alpha)}{2-\alpha} \right]^3 \left(\frac{2}{75} \times \frac{2-\alpha}{2} \right) + \frac{12(1-\alpha)^2\alpha}{(2-\alpha)^3} \left(\frac{1}{60} \times \frac{2-\alpha}{2} \right), \\
 &= \frac{[2(1-\alpha)]^3}{60(2-\alpha)^2} + \frac{\alpha}{30} \left[\frac{2(1-\alpha)}{2-\alpha} \right]^2.
 \end{aligned}$$

g) Enfin, en approximant la dernière somme de l'équation (3.2.6), on a

$$\begin{aligned}
& \lim_{N \rightarrow \infty} \sum_{t \geq 0} E_0(Q_{AB}^3(t)) \\
&= P_{[15,12]} \left(\mu_{25} \int_0^\infty (e^{PRP\tau})_{[12,6]} d\tau + \mu_{26} \int_0^\infty (e^{PRP\tau})_{[12,4]} d\tau + \mu_{27} \int_0^\infty (e^{PRP\tau})_{[12,2]} d\tau \right) .
\end{aligned} \tag{3.2.18}$$

Il y a ici trois possibilités de chemins qui constituent la probabilité qu'il y ait trois individus de génotype AB au temps τ sachant qu'il n'y a qu'un allèle A au temps 0. Dans tous ces cas, puisqu'on doit remonter les lignées ancestrales des deux gènes de trois individus au temps τ , la configuration de départ de ces gènes est représenté par l'état 15. Puisque les deux gènes de chaque individu sont de type différent au temps τ , il est impossible qu'il y ait coalescence instantanée, donc la seule possibilité de transition instantanée est celle de l'état 15 à l'état 12 ($P_{[15,12]}$). Suite à ce passage instantané, les six gènes ancestraux A, A, A, B, B, B sont dans six individus différents et ce, dans les trois possibilités de chemins.

Le premier terme dans la parenthèse de l'équation (3.2.18) représente un lent passage de l'état 12, avec A, A, A, B, B, B , à l'état 6, avec A, B, B, B au temps 0 (d'où $(e^{PRP\tau})_{[12,6]}$). Puisque le type des gènes ancestraux importe lors de ce passage, il faut multiplier par la probabilité que les paires choisies pour coalescer soient toutes de type A . Cette probabilité est $\mu_{25} = 3/150$. Le deuxième cas (représenté par le deuxième terme dans la parenthèse du membre de droite de l'équation (3.2.18)), est la probabilité d'un passage de l'état 12 à l'état 4 (d'où $(e^{PRP\tau})_{[12,4]}$). Il y a alors trois coalescences lors du processus lent. Il faut donc multiplier par la probabilité que le passage de A, A, A, B, B, B à A, B, B soit compatible avec les trois coalescences. On trouve que cette probabilité est $\mu_{26} = 3/100$. Le dernier terme dans la parenthèse de l'équation (3.2.18) correspond à un lent passage de l'état 12, avec A, A, A, B, B, B , à l'état 2, avec A, B , (d'où $(e^{PRP\tau})_{[12,2]}$) en passant par quatre coalescences. Puisque ces coalescences doivent se produire entre des gènes ancestraux de même type, on multiplie donc par la probabilité de choisir une paire de gène de même type parmi toutes les autres

paires possibles, et ce, avant chaque coalescence. Cette probabilité est $\mu_{27} = 1/50$.

En évaluant la limite, on obtient

$$\begin{aligned}
 & \left[\frac{2(1-\alpha)}{2-\alpha} \right]^3 \left[\frac{3}{150} \times \int_0^\infty \frac{5 e^{-12\tau/(2-\alpha)}}{6} \left(-9 e^{-8\tau/(2-\alpha)} + 4 e^{-18\tau/(2-\alpha)} + 5 \right) d\tau \right. \\
 & \quad \left. + \left(\frac{3}{100} \times \frac{2-\alpha}{6} \right) + \left(\frac{1}{50} \times \frac{2-\alpha}{2} \right) \right], \\
 &= \left[\frac{2(1-\alpha)}{2-\alpha} \right]^3 \left[\left(\frac{3}{150} \times \frac{2-\alpha}{12} \right) + \left(\frac{3}{100} \times \frac{2-\alpha}{6} \right) + \left(\frac{1}{50} \times \frac{2-\alpha}{2} \right) \right], \\
 &= \frac{[2(1-\alpha)]^3}{60(2-\alpha)^2}.
 \end{aligned}$$

En revenant finalement à l'équation (3.2.6) et en remplaçant les sommes d'espérances par leurs limites, on a

$$\begin{aligned}
 & \lim_{N \rightarrow \infty} F'_A(0) \\
 &= a_3 \left(\frac{\alpha^2}{2(2-\alpha)} + \frac{4\alpha(1-\alpha)}{3(2-\alpha)} + \frac{[2(1-\alpha)]^2}{6(2-\alpha)} \right) + b_3(1-\alpha) \\
 & \quad + c_3 \left(\frac{[2(1-\alpha)]^3}{20(2-\alpha)^2} + \frac{13\alpha}{60} \left[\frac{2(1-\alpha)}{2-\alpha} \right]^2 + \frac{2(1-\alpha)}{3} \left[\frac{\alpha}{2-\alpha} \right]^2 + \frac{\alpha^3}{6(2-\alpha)^2} \right) \\
 & \quad + d_3 \left(\frac{\alpha(1-\alpha)}{3(2-\alpha)} + \frac{[2(1-\alpha)]^2}{12(2-\alpha)} \right) + e_3 \left(\frac{[2(1-\alpha)]^2}{12(2-\alpha)} \right) \\
 & \quad + g_3 \left(\frac{[2(1-\alpha)]^3}{30(2-\alpha)^2} + \frac{\alpha[2(1-\alpha)]^2}{10(2-\alpha)^2} + \frac{\alpha^2(1-\alpha)}{6(2-\alpha)^2} \right) \\
 & \quad + h_3 \left(\frac{[2(1-\alpha)]^3}{60(2-\alpha)^2} + \frac{\alpha}{30} \left[\frac{2(1-\alpha)}{2-\alpha} \right]^2 \right) + k_3 \left(\frac{[2(1-\alpha)]^3}{60(2-\alpha)^2} \right).
 \end{aligned}$$

En réorganisant l'équation par rapport aux termes $(\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB})$ et $(\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M\mathbf{p}_{BB}$, on trouve

$$\begin{aligned} \lim_{N \rightarrow \infty} F'_A(0) = & \left(\frac{86 - 52\alpha - 5\alpha^2 + 12\alpha^3 - \alpha^4}{240(2 - \alpha)^2} \right) (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB}) \\ & + \left(\frac{5 - \alpha}{8(2 - \alpha)} \right) (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M\mathbf{p}_{BB} . \end{aligned} \quad (3.2.19)$$

Finalement, on obtient l'approximation

$$\begin{aligned} F_A(s) = & \frac{1}{2N} + s \left[\left(\frac{86 - 52\alpha - 5\alpha^2 + 12\alpha^3 - \alpha^4}{240(2 - \alpha)^2} \right) (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M(\mathbf{p}_{AA} - \mathbf{p}_{BB}) \right. \\ & \left. + \left(\frac{5 - \alpha}{8(2 - \alpha)} \right) (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet M\mathbf{p}_{BB} \right] + o(s) . \end{aligned} \quad (3.2.20)$$

en supposant que s est d'ordre $1/N$.

3.2.2. Analyse de la probabilité de fixation

On peut analyser cette probabilité de fixation de manière à trouver une stratégie optimale. On sait que sans sélection, la probabilité de fixation d'un allèle est égale à la fréquence initiale de celui-ci (voir le chapitre 2.3). On peut donc comparer la probabilité de fixation lorsqu'il y a sélection avec celle du modèle neutre pour savoir si elle augmente ou diminue. La probabilité de fixation dans notre cas avec sélection peut s'exprimer ainsi :

$$F_A(x_0, s) = x_0 + s \times \Gamma(\nu_{AA}, \nu_{BB}) + o(s) ,$$

où ν_{ij} est la probabilité qu'un individu ij choisisse d'utiliser la stratégie 1, c'est-à-dire, $\mathbf{p}_{ij} = (\nu_{ij}, 1 - \nu_{ij})$. Pour que la stratégie \mathbf{p}_{BB} soit une ESS, il faut qu'elle

respecte les conditions d'équilibre (voir l'inégalité (3.1.2)) et de stabilité (voir l'inégalité (3.1.3)). On rappelle qu'on travaille avec une population finie dans laquelle la fréquence initiale des gènes de type A est $1/2N$ et celle des gènes de type B est $1 - 1/2N$. On a soulevé, dans la section 3.1, qu'avec une population finie, on peut trouver une stratégie optimale à partir de la probabilité de fixation et la comparer à une stratégie stable au cours de l'évolution (ESS). On doit cependant adapter les conditions d'équilibre et de stabilité. D'abord, puisque sans sélection, $F_A(x_0) = x_0$, et que la stratégie \mathbf{p}_{BB} doit être avantagée sélectivement par rapport à la stratégie \mathbf{p}_{AA} (pour que la stratégie \mathbf{p}_{BB} soit optimale), il faut que $F_A(x_0) \leq x_0$ et donc que $\Gamma(\nu_{AA}, \nu_{BB}) \leq 0$. Ensuite, on doit avoir que

$$\left[\frac{\partial}{\partial \nu_{AA}} \Gamma(\nu_{AA}, \nu_{BB}) \right]_{\nu_{AA}=\nu_{BB}} = 0 ,$$

pour l'équilibre, et aussi que

$$\left[\frac{\partial^2}{\partial \nu_{AA}^2} \Gamma(\nu_{AA}, \nu_{BB}) \right]_{\nu_{AA}=\nu_{BB}} < 0 ,$$

pour la stabilité, en faisant fit des cas de dégénérescence. Ces deux dernières conditions signifient que la probabilité de fixation de A atteint un maximum local lorsque $\mathbf{p}_{AA} = \mathbf{p}_{BB}$ et donc que la stratégie \mathbf{p}_{BB} est la meilleure stratégie à adopter contre elle-même. Ces conditions signifient aussi que \mathbf{p}_{BB} est la meilleure stratégie à adopter contre une stratégie \mathbf{p}_{AA} telle que \mathbf{p}_{AA} est assez près, mais différente de \mathbf{p}_{BB} . Ces conditions sont donc équivalentes aux conditions (3.1.2) et (3.1.3). En exprimant la matrice de jeux sous la forme

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} ,$$

et en utilisant l'équation (3.2.20), on a

$$\begin{aligned}
& \Gamma(\nu_{AA}, \nu_{BB}) \\
&= \left(\frac{86 - 52\alpha - 5\alpha^2 + 12\alpha^3 - \alpha^4}{240(2 - \alpha)^2} \right) (\nu_{AA} - \nu_{BB})^2 (m_{11} - m_{21} - m_{12} + m_{22}) \\
&+ \left(\frac{5 - \alpha}{8(2 - \alpha)} \right) (\nu_{AA} - \nu_{BB}) [\nu_{BB}(m_{11} - m_{12}) + (1 - \nu_{BB})(m_{12} - m_{22})] .
\end{aligned}$$

On trouve donc

$$\begin{aligned}
& \left[\frac{\partial}{\partial \nu_{AA}} \Gamma(\nu_{AA}, \nu_{BB}) \right]_{\nu_{AA}=\nu_{BB}} \\
&= \left(\frac{5 - \alpha}{8(2 - \alpha)} \right) [\nu_{BB}(m_{11} - m_{21} - m_{12} + m_{22}) + m_{12} - m_{22}] ,
\end{aligned}$$

qui est égale à 0, pour $0 < \alpha < 1$, si et seulement si

$$\nu_{BB} = \frac{-m_{12} + m_{22}}{m_{11} - m_{21} - m_{12} + m_{22}} .$$

Pour la dérivée seconde, on trouve

$$\begin{aligned}
& \left[\frac{\partial^2}{\partial \nu_{AA}^2} \Gamma(\nu_{AA}, \nu_{BB}) \right]_{\nu_{AA}=\nu_{BB}} \\
&= 2 \left(\frac{86 - 52\alpha - 5\alpha^2 + 12\alpha^3 - \alpha^4}{240(2 - \alpha)^2} \right) (m_{11} - m_{21} - m_{12} + m_{22}) ,
\end{aligned}$$

qui est négative si et seulement si

$$m_{11} - m_{21} - m_{12} + m_{22} < 0 ,$$

car, puisque $0 < \alpha < 1$, on a que

$$86 - 52\alpha - 5\alpha^2 + 12\alpha^3 - \alpha^4 > 0 .$$

En combinant le tout, si $m_{11} - m_{21} - m_{12} + m_{22} < 0$, alors

$$\mathbf{p}_{BB} = \left(\frac{-m_{12} + m_{22}}{m_{11} - m_{21} - m_{12} + m_{22}}, \frac{m_{11} - m_{21}}{m_{11} - m_{21} - m_{12} + m_{22}} \right)$$

est une stratégie optimale. Mais puisque les composantes de \mathbf{p}_{BB} doivent être entre 0 et 1, il faut donc que $-m_{12} + m_{22} < 0$ (autrement on pose la première composante de $\mathbf{p}_{BB} = 0$) et que $m_{11} - m_{21} < 0$ (autrement on pose la deuxième composante de $\mathbf{p}_{BB} = 0$). On remarque que α n'entre pas dans la composition du vecteur \mathbf{p}_{BB} et qu'alors, un changement dans la probabilité d'autofécondation ne change pas la valeur de la stratégie optimale.

3.2.3. Probabilité de fixation dans une population avec autofécondation partielle sans interaction

Calculons maintenant la probabilité de fixation pour un cas particulier du dernier modèle : on suppose qu'il y a encore de l'autofécondation partielle mais pas d'interaction.

Le calcul de cette probabilité de fixation est un peu plus simple. En considérant le même cycle que celui de la page 50, la même valeur initiale de la fréquence de A

($x_0 = 1/2N$), la même méthode que celle employée précédemment et les valeurs sélectives

$$w_{AA}(t) = 1 + s u_{AA} ,$$

$$w_{AB}(t) = 1 + s u_{AB} ,$$

$$w_{BB}(t) = 1 + s u_{BB} ,$$

car il n'y a pas d'interaction, où

$$u_{AB} = \frac{u_{AA} + u_{BB}}{2} ,$$

et donc $w_{AB} = (w_{AA} + w_{BB})/2$, on trouve

$$\begin{aligned} F'_A(0) = & \left[\left(\frac{u_{AA} - u_{BB}}{2} \right) (1 + \alpha) \right] \sum_{t \geq 0} E_0 \left(Q_{AA}(t) - Q_{AA}^2(t) \right) \\ & + \left[\left(\frac{u_{AA} - u_{BB}}{4} \right) \left(1 + \frac{\alpha}{2} \right) \right] \sum_{t \geq 0} E_0 \left(Q_{AB}(t) \right) \\ & + \left[\left(\frac{u_{BB} - u_{AA}}{2} \right) (1 + \alpha) \right] \sum_{t \geq 0} E_0 \left(Q_{AA}(t) Q_{BB}(t) \right) \\ & + \left[\left(\frac{u_{BB} - u_{AA}}{8} \right) (1 + \alpha) \right] \sum_{t \geq 0} E_0 \left(Q_{AB}^2(t) \right) . \end{aligned}$$

Pour calculer les sommes d'espérances ci-dessus, on utilise le lemme C.0.1 (de Möhle) et on a besoin de la matrice Π qui rassemblent les probabilités de transition des configurations possibles de gènes ancestraux d'une génération à la précédente. Dans ce cas cependant, on a seulement à considérer huit états (et non quinze), car on a quatre gènes ancestraux à suivre au maximum (et non six). On travaille donc avec une sous-matrice de la matrice Π considérée précédemment dans le cas avec interaction. En effet, dans les deux cas, on a de l'autofécondation partielle et

on veut les probabilités de transition lorsqu'il n'y a pas de sélection ($s = 0$), donc les entrées correspondantes des deux matrices seront les mêmes. Il s'en suit que les matrices $H, P, R_N, R, e^{PRP\tau}$ ont les mêmes éléments que ceux des matrices correspondantes dans le cas avec interaction et donc que l'approximation des sommes d'espérances ci-dessus ont la même valeur que précédemment. Il s'ensuit que

$$\begin{aligned} \lim_{N \rightarrow \infty} F'_A(0) &= \left[\left(\frac{u_{AA} - u_{BB}}{2} \right) (1 + \alpha) \right] \frac{4 - \alpha^2}{6(2 - \alpha)} \\ &\quad + \left[\left(\frac{u_{AA} - u_{BB}}{4} \right) \left(1 + \frac{\alpha}{2} \right) \right] (1 - \alpha) \\ &\quad + \left[\left(\frac{u_{BB} - u_{AA}}{2} \right) (1 + \alpha) \right] \frac{1 - \alpha}{3(2 - \alpha)} \\ &\quad + \left[\left(\frac{u_{BB} - u_{AA}}{8} \right) (1 + \alpha) \right] \frac{1 - 2\alpha + \alpha^2}{3(2 - \alpha)} , \\ &= \frac{5 - \alpha}{8(2 - \alpha)} (u_{AA} - u_{BB}) . \end{aligned}$$

Donc la probabilité de fixation de A dans un modèle additif d'une population diploïde avec autofécondation partielle mais sans interaction est approximée par

$$F_A(s) = \frac{1}{2N} + s \left(\frac{5 - \alpha}{8(2 - \alpha)} (u_{AA} - u_{BB}) \right) + o(s) ,$$

dans le cas où l'intensité de sélection s est d'ordre $1/N$ et N est grand. Comme il n'y a pas de stratégie dans ce modèle, on ne peut évidemment pas trouver de stratégie stable au cours de l'évolution. Cependant, on peut observer que la probabilité de fixation est plus petite que la fréquence initiale de A seulement si $u_{BB} > u_{AA}$. Donc, plus les individus de type BB sont avantagés, plus la probabilité d'extinction de A est élevée.

3.2.4. Probabilité de fixation dans un modèle de fertilité

Considérons maintenant un modèle dans lequel c'est la fertilité, contrairement à la viabilité, qui diffère selon le génotype. Le cycle d'un tel modèle n'est pas le même que celui de la page 50, car la sélection vient avant la reproduction. On considère une population diploïde de taille N (donc de $2N$ gènes) avec autofécondation partielle (la probabilité d'autofécondation est α). On veut encore la probabilité éventuelle de fixation de A sachant qu'à la génération initiale la fréquence de A est $1/2N$ ($X(0) = 1/2N$). On suppose, pour la fertilité, un grand nombre d'affrontements entre les individus adultes avant la reproduction donc, comme un individu ne peut s'affronter lui-même, on a les valeurs sélectives suivantes :

$$\begin{aligned} w_{AA}(t) &= 1 + s\mathbf{p}_{AA} \bullet M\left(\bar{\mathbf{p}}(t) - \frac{\mathbf{p}_{AA}}{N}\right), \\ w_{AB}(t) &= 1 + s\mathbf{p}_{AB} \bullet M\left(\bar{\mathbf{p}}(t) - \frac{\mathbf{p}_{AB}}{N}\right), \\ w_{BB}(t) &= 1 + s\mathbf{p}_{BB} \bullet M\left(\bar{\mathbf{p}}(t) - \frac{\mathbf{p}_{BB}}{N}\right), \end{aligned}$$

où $s = d/(N-1)$ et d est petit, $\mathbf{p}_{AB} = (\mathbf{p}_{BB} + \mathbf{p}_{AA})/2$, et donc

$$\bar{\mathbf{p}}(t) = X(t)(\mathbf{p}_{AA} - \mathbf{p}_{BB}) + \mathbf{p}_{BB}.$$

On utilise la même méthode que précédemment, mais en inversant l'ordre de la sélection et de la reproduction. On a encore que

$$F_A(s) = \frac{1}{2N} + sF'_A(0) + o(s),$$

où

$$F'_A(0) = \sum_{t=0}^{\infty} \sum_{\mathbf{q}} \left[\left(\frac{d}{ds} E_s \left(X(t+1) - X(t) \mid \mathbf{Q}(t) = \mathbf{q} \right) \right) \Big|_{s=0} \times \text{Pr}_0(\mathbf{Q}(t) = \mathbf{q}) \right],$$

avec $\mathbf{Q}(t) = (Q_{AA}(t), Q_{AB}(t), Q_{BB}(t))$ et $\mathbf{q} = (q_{AA}, q_{AB}, q_{BB})$. On trouve que

$$\begin{aligned} E_s \left(X(t+1) - X(t) \mid \mathbf{Q}(t) = \mathbf{q} \right) = & [a_4 \mathbf{p}_{AA} \bullet M \mathbf{p}_{AA} + b_4 \mathbf{p}_{AA} \bullet M \mathbf{p}_{BB} \\ & + c_4 \mathbf{p}_{BB} \bullet M \mathbf{p}_{AA} + d_4 \mathbf{p}_{BB} \bullet M \mathbf{p}_{BB}] s + o(s), \end{aligned} \quad (3.2.21)$$

où

$$\begin{aligned} a_4 = & -\frac{q_{AA}}{N} + \left(1 + \frac{1}{N}\right) q_{AA}^2 - q_{AA}^3 - \frac{q_{AB}}{8N} + \frac{3}{4} \left(1 + \frac{1}{N}\right) q_{AA} q_{AB} - \frac{3q_{AA}^2 q_{AB}}{2} \\ & + \frac{1}{8} \left(1 + \frac{1}{N}\right) q_{AB}^2 - \frac{3q_{AA} q_{AB}^2}{4} - \frac{q_{AB}^3}{8}, \end{aligned}$$

$$\begin{aligned} b_4 = & q_{AA} - 2q_{AA}^2 + q_{AA}^3 + \frac{1}{4} \left(1 - \frac{1}{2N}\right) q_{AB} - \left(\frac{7}{4} - \frac{1}{4N}\right) q_{AA} q_{AB} + \frac{3q_{AA}^2 q_{AB}}{2} \\ & - \left(\frac{3}{8} - \frac{1}{8N}\right) q_{AB}^2 + \frac{3q_{AA} q_{AB}^2}{4} + \frac{q_{AB}^3}{8}, \end{aligned}$$

$$\begin{aligned} c_4 = & -q_{AA}^2 + q_{AA}^3 - \frac{q_{AB}}{8N} - \left(\frac{3}{4} - \frac{1}{4N}\right) q_{AA} q_{AB} + \frac{3q_{AA}^2 q_{AB}}{2} - \frac{1}{8} \left(1 - \frac{1}{N}\right) q_{AB}^2 \\ & + \frac{3q_{AA} q_{AB}^2}{4} + \frac{q_{AB}^3}{8}, \end{aligned}$$

$$\begin{aligned} d_4 = & -\left(1 - \frac{1}{N}\right) q_{AA} + \left(2 - \frac{1}{N}\right) q_{AA}^2 - q_{AA}^3 - \left(\frac{1}{4} - \frac{3}{8N}\right) q_{AB} + \left(\frac{7}{4} - \frac{5}{4N}\right) q_{AA} q_{AB} \\ & - \frac{3q_{AA}^2 q_{AB}}{2} + \frac{3}{8} \left(1 - \frac{1}{N}\right) q_{AB}^2 - \frac{3q_{AA} q_{AB}^2}{4} - \frac{q_{AB}^3}{8}. \end{aligned}$$

En dérivant l'équation (3.2.21) par rapport à s et en évaluant à $s = 0$, on obtient

$$\begin{aligned} \frac{d}{ds} \left[E_s \left(X(t+1) - X(t) \mid \mathbf{Q}(t) = \mathbf{q} \right) \right]_{s=0} &= a_4 \mathbf{p}_{AA} \bullet M \mathbf{p}_{AA} + b_4 \mathbf{p}_{AA} \bullet M \mathbf{p}_{BB} \\ &\quad + c_4 \mathbf{p}_{BB} \bullet M \mathbf{p}_{AA} + d_4 \mathbf{p}_{BB} \bullet M \mathbf{p}_{BB} . \end{aligned}$$

Finalement, en multipliant par $Pr_0((\mathbf{Q}(t) = \mathbf{q}))$ et en sommant sur t et \mathbf{q} , on trouve

$$\begin{aligned} F'_A(0) &= a_5 \sum_{t \geq 0} E_0(Q_{AA}(t) - Q_{AA}^2(t)) + b_5 \sum_{t \geq 0} E_0(Q_{AA}^2(t) - Q_{AA}^3(t)) + c_5 \sum_{t \geq 0} E_0(Q_{AB}(t)) \\ &\quad + d_5 \sum_{t \geq 0} E_0(Q_{AA}(t)Q_{AB}(t)) + e_5 \sum_{t \geq 0} E_0(Q_{AB}^2(t)) + g_5 \sum_{t \geq 0} E_0(Q_{AA}^2(t)Q_{AB}(t)) \\ &\quad + h_5 \sum_{t \geq 0} E_0(Q_{AA}(t)Q_{AB}^2(t)) + k_5 \sum_{t \geq 0} E_0(Q_{AB}^3(t)) , \end{aligned} \quad (3.2.22)$$

où

$$a_5 = -\frac{1}{N}(\mathbf{p}_{AA} \bullet M \mathbf{p}_{AA}) + (\mathbf{p}_{AA} \bullet M \mathbf{p}_{BB}) - \left(1 - \frac{1}{N}\right)(\mathbf{p}_{BB} \bullet M \mathbf{p}_{BB}) ,$$

$$b_5 = (\mathbf{p}_{AA} \bullet M \mathbf{p}_{AA}) - (\mathbf{p}_{AA} \bullet M \mathbf{p}_{BB}) - (\mathbf{p}_{BB} \bullet M \mathbf{p}_{AA}) + (\mathbf{p}_{BB} \bullet M \mathbf{p}_{BB}) ,$$

$$\begin{aligned} c_5 &= -\frac{1}{8N}(\mathbf{p}_{AA} \bullet M \mathbf{p}_{AA}) + \frac{1}{4}\left(1 - \frac{1}{2N}\right)(\mathbf{p}_{AA} \bullet M \mathbf{p}_{BB}) - \frac{1}{8N}(\mathbf{p}_{BB} \bullet M \mathbf{p}_{AA}) \\ &\quad - \left(\frac{1}{4} - \frac{3}{8N}\right)(\mathbf{p}_{BB} \bullet M \mathbf{p}_{BB}) , \end{aligned}$$

$$d_5 = \frac{3}{4} \left(1 + \frac{1}{N} \right) (\mathbf{p}_{AA} \bullet M \mathbf{p}_{AA}) - \left(\frac{7}{4} - \frac{1}{4N} \right) (\mathbf{p}_{AA} \bullet M \mathbf{p}_{BB}) \\ - \left(\frac{3}{4} - \frac{1}{4N} \right) (\mathbf{p}_{BB} \bullet M \mathbf{p}_{AA}) + \left(\frac{7}{4} - \frac{5}{4N} \right) (\mathbf{p}_{BB} \bullet M \mathbf{p}_{BB}) ,$$

$$e_5 = \frac{1}{8} \left(1 + \frac{1}{N} \right) (\mathbf{p}_{AA} \bullet M \mathbf{p}_{AA}) - \left(\frac{3}{8} - \frac{1}{8N} \right) (\mathbf{p}_{AA} \bullet M \mathbf{p}_{BB}) \\ - \frac{1}{8} \left(1 - \frac{1}{N} \right) (\mathbf{p}_{BB} \bullet M \mathbf{p}_{AA}) + \frac{3}{8} \left(1 - \frac{1}{N} \right) (\mathbf{p}_{BB} \bullet M \mathbf{p}_{BB}) ,$$

$$g_5 = -\frac{3}{2} (\mathbf{p}_{AA} \bullet M \mathbf{p}_{AA}) + \frac{3}{2} (\mathbf{p}_{AA} \bullet M \mathbf{p}_{BB}) + \frac{3}{2} (\mathbf{p}_{BB} \bullet M \mathbf{p}_{AA}) - \frac{3}{2} (\mathbf{p}_{BB} \bullet M \mathbf{p}_{BB}) ,$$

$$h_5 = -\frac{3}{4} (\mathbf{p}_{AA} \bullet M \mathbf{p}_{AA}) + \frac{3}{4} (\mathbf{p}_{AA} \bullet M \mathbf{p}_{BB}) + \frac{3}{4} (\mathbf{p}_{BB} \bullet M \mathbf{p}_{AA}) - \frac{3}{4} (\mathbf{p}_{BB} \bullet M \mathbf{p}_{BB}) ,$$

$$k_5 = -\frac{1}{8} (\mathbf{p}_{AA} \bullet M \mathbf{p}_{AA}) + \frac{1}{8} (\mathbf{p}_{AA} \bullet M \mathbf{p}_{BB}) + \frac{1}{8} (\mathbf{p}_{BB} \bullet M \mathbf{p}_{AA}) - \frac{1}{8} (\mathbf{p}_{BB} \bullet M \mathbf{p}_{BB}) .$$

Ces sommes d'espérances peuvent être approximées de la même façon qu'à la section 3.2.1 avec la théorie de la coalescence et le lemme C.0.1 de Möhle. Le fait qu'on travaille ici avec un modèle de fertilité ne change pas la précédente valeur trouvée des sommes de l'équation (3.2.22), car on approxime celles-ci lorsque $s = 0$. Les éléments et la taille des matrices H, P, R_N, R , et $e^{PRP\tau}$ du modèle de fertilité sont donc les mêmes que ceux du modèle de viabilité. En remplaçant les sommes d'espérance par leur limite, on trouve

$$\begin{aligned}
F'_A(0) = & \frac{5(-30 + 17\alpha + 3\alpha^2 - 2\alpha^3) + N(56 - 53\alpha + 18\alpha^2 - \alpha^3)}{120 N(2 - \alpha)^2} \mathbf{p}_{AA} \bullet M \mathbf{p}_{AA} \\
& + \frac{5(-6 + 13\alpha - 9\alpha^2 + 2\alpha^3) + N(124 - 97\alpha + 12\alpha^2 + \alpha^3)}{120 N(2 - \alpha)^2} \mathbf{p}_{AA} \bullet M \mathbf{p}_{BB} \\
& + \frac{5(-6 + 13\alpha - 9\alpha^2 + 2\alpha^3) - N(56 - 53\alpha + 18\alpha^2 - \alpha^3)}{120 N(2 - \alpha)^2} \mathbf{p}_{BB} \bullet M \mathbf{p}_{AA} \\
& + \frac{5(42 - 43\alpha + 15\alpha^2 - 2\alpha^3) - N(124 - 97\alpha + 12\alpha^2 + \alpha^3)}{120 N(2 - \alpha)^2} \mathbf{p}_{BB} \bullet M \mathbf{p}_{BB} .
\end{aligned}$$

En prenant la matrice $\tilde{M} = M - \frac{1}{N}(M + M^T)$, on obtient que

$$\begin{aligned}
F_A(s) = & \frac{1}{2N} + s \left[K (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet \tilde{M} (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \right. \\
& \left. + \frac{3 - \alpha}{4(2 - \alpha)} (\mathbf{p}_{AA} - \mathbf{p}_{BB}) \bullet \tilde{M} \mathbf{p}_{BB} \right] + o(s) , \quad (3.2.23)
\end{aligned}$$

où

$$K = \frac{5(-30 + 17\alpha + 3\alpha^2 - 2\alpha^3) + N(56 - 53\alpha + 18\alpha^2 - \alpha^3)}{120 (-2 + N)(2 - \alpha)^2} .$$

La forme de l'équation (3.2.23) est la même que celle de l'équation (3.2.20) si on considère N grand. Dans ce cas, les conclusions pour trouver une stratégie optimale sont semblables.

CONCLUSION

On a vu plusieurs façons de calculer la probabilité de fixation d'un nouveau gène mutant d'une population diploïde de taille $2N$. Premièrement, dans la section 2.2, on a présenté une méthode approximative par le processus de diffusion sur un modèle avec sélection seulement. Pour utiliser cette méthode, on doit prendre $2N$ générations comme unité de temps et supposer une population de taille très grande. Cette méthode est employée souvent et il est possible d'en voir d'autres exemples dans [Le(2)]. Le résultat de cette dernière a ensuite été comparé à celui de la méthode exacte à la section 2.3 pour une population de taille arbitraire. On a pu observer leur similitude lorsque N est grand. La méthode exacte, utilisée avec une population de taille grande mais finie, est plus récente. L'expansion de Taylor, par rapport à l'intensité de la sélection s , de la probabilité de fixation est employée, en supposant la dérivabilité et la continuité en s de celle-ci. Le désavantage de cette méthode est la complexité du calcul des espérances des temps de coalescence qui apparaissent dans le coefficient de s . Pour des modèles simples comme celui de la section 2.3, on a seulement l'espérance des temps de coalescence de deux et trois gènes à calculer. Mais plus les modèles sont complexes, plus le nombre de gènes dans le calcul des temps de coalescence est élevé, et plus le calcul est compliqué.

Au chapitre trois, on a d'abord vu un modèle de viabilité avec interaction entre les individus et autofécondation partielle. Ce modèle étant plus complexe que les précédents, la méthode exacte n'a pas été employée. On a quand même utilisé l'expansion de Taylor pour la probabilité de fixation, mais c'est dans le calcul des termes de l'expansion qu'on a dû changer la manière de procéder. On

s'est encore servi de la théorie de la coalescence, mais cette fois, on a dû exprimer la probabilité de fixation autrement et utiliser le lemme de Möhle ([M]) pour calculer la somme sur t de la probabilité du génotype d'un individu de la génération t . Ce calcul aussi est compliqué et il serait bien de le comparer à celui de la méthode exacte pour différents modèles.

Après avoir trouvé la probabilité de fixation pour ce dernier modèle, on a pu analyser celle-ci pour voir si une stratégie stable au cours de l'évolution (ESS) est optimale par rapport à la probabilité de fixation. Les valeurs de la stratégie optimale trouvée changent avec les valeurs de la matrice de paiement M , car cette matrice donne les avantages sélectifs correspondants aux différentes stratégies employées. Cette stratégie ne dépend cependant pas de α , le coefficient d'autofécondation partielle et correspond à un ESS. On a par la suite calculé la probabilité de fixation dans le modèle de fertilité avec interaction et autofécondation partielle. On a vu que les mêmes conclusions que celles du cas précédent se présentaient si on posait $\tilde{M} = M - \frac{1}{N}(M + M^T)$.

On a donc étudié l'intensité de l'avantage sélectif qu'un nouveau gène mutant doit avoir pour rester dans la population et éliminer les autres gènes moins avantageux. On s'en est servi pour trouver des stratégies optimales et donc un peu mieux comprendre l'évolution de certains modèles de population. Dans le futur, il serait intéressant de calculer la probabilité de fixation de plusieurs autres modèles diploïdes. Le modèle de Moran ainsi que celui avec des sous-populations et de la migration pourraient être envisagés.

BIBLIOGRAPHIE

- [BBC] EDITED BY D.J. BALDING, M. BISHOP, C. CANNINGS, *Handbook of Statistical Genetics (Coalescent Theory*, M. Nordborg), Wiley, Chichester (UK), 2001.
- [Be] http://www.scs.fsu.edu/~beerli/BSC-5936/10-31-05/lecture_16.pdf , site internet de Peter Beerli à l'Université de l'état de la Floride.
- [Bi] P. BILLINGSLEY, *Probability and Measure*, John Wiley & Sons Inc., New York (USA), 1995.
- [Bu] E. BURGER, *Introduction to the Theory of Games*, Prentice-Hall INC., New Jersey (USA), 1963.
- [D] C. DARWIN, *On the Origin of Species*, John Murray, Londre, 1859.
- [Da] M. D. DAVIS, *Game Theory, a Nontechnical Introduction*, Basic Books INC., New York (USA), 1970.
- [Du] R. DURRETT, *Essentials of Stochastic Processes*, Springer, New York, 1999.
- [E] W.J. EWENS, *Mathematical Population Genetics, 1. Theoretical Introduction*, Springer-Verlag, New York, 2004.
- [F] R.A. FISHER, *The Genetical Theory of Natural Selection*, Clarendon Press, Oxford, 1930.
- [HS] J. HOFBAUER, K. SIGMUND, *Evolutionary Games and Population Dynamics*, Cambridge University Press, Cambridge, 1998.
- [KT] S. KARLIN, H. M. TAYLOR, *A Second Course in Stochastic Processes*, Academic Press, New York, 1981.
- [K] M. KIMURA, *On the probability of fixation of mutant genes in a population*, 1962. *Genetics* 47 : 713-719.

- [Ki(a)] J.F.C. KINGMAN, *The coalescent*, 1982. Stochastic Processes Applications 13 : 235-248.
- [Ki(b)] J.F.C. KINGMAN, *On the genealogy of large populations*, 1982. Journal of Applied Probability 19A : 27-43.
- [L] J.B. LAMARCK, *Philosophie Zoologique*, Dentu, Paris, 1809.
- [La] K. LANGE, *Mathematical and Statistical Methods for Genetic Analysis*, Springer, New York, 2002.
- [Le(1)] S. LESSARD, *The theory of optimal sex ratio*, 2002. Comments on Theoretical Biology 7 : 315-354.
- [Le(2)] S. LESSARD, *Long-term stability from fixation probabilities in finite populations : New perspectives for ESS theory*, 2005. Theoretical Population Biology 68 : 19-27.
- [LL] S. LESSARD, V. LADRET, *The probability of fixation of a single mutant in an exchangeable selection model*, article soumis.
- [M] M. MÖHLE, *A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing*, 1998. Advances in Applied Probability 30 : 493-512.
- [T] S. TAVARE, O. ZEITOUNI, *Lectures on Probability Theory and Statistics*, Springer, Berlin, 2004.
- [VM] J. VON NEUMANN, O. MORGENSTERN *Theory of Games and Economic Behavior*, Wiley, New York, 1944.
- [W] S. WRIGHT, *Evolution in Mendelian populations*, 1931. Genetics 16 : 97-159.

Annexe A

CALCUL EXACT DE L'ESPÉRANCE DE T_3 LORSQUE $S = 0$

On suppose une population diploïde de N individus (donc $2N$ gènes) qui suit le modèle de Wright-Fisher. Soit T_3 le nombre de générations qu'il faut remonter avant d'arriver à l'ancêtre commun le plus récent (MRCA) de trois gènes. On veut trouver l'espérance exacte de T_3 lorsqu'il n'y a pas de sélection.

Soit β_{ij} la probabilité de passer de i à j gènes ancestraux en une génération en remontant le temps ($i \geq j$) lorsqu'il n'y a pas de sélection. Avec trois gènes ancestraux à suivre au maximum, on utilisera les probabilités β_{33} , β_{32} , β_{31} , β_{22} et β_{21} pour trouver l'espérance de T_3 . On a vu à la section 1.2 que $\beta_{21} = 1/2N$ et que $\beta_{22} = 1 - \beta_{21} = 1 - 1/2N$. De la même façon, on a

$$\beta_{31} = 1/4N^2 ,$$

car la probabilité que le gène parental du second gène ancestral soit le même que celui du premier est $1/2N$. Puisque la probabilité que le troisième gène ancestral ait aussi le même parent est $1/2N$, la probabilité que les trois gènes ancestraux aient le même parent est donc $1/2N \times 1/2N = 1/4N^2$. Ensuite, pour passer de 3 à 2 gènes ancestraux en une génération en remontant le temps, il faut qu'exactement deux des trois gènes ancestraux aient le même gène parental. Puisqu'on doit choisir deux gènes ancestraux qui vont coalescer parmi trois, on a

$$\beta_{32} = \binom{3}{2} \times \frac{1}{2N} \left(1 - \frac{1}{2N}\right) = \frac{3}{2N} \left(1 - \frac{1}{2N}\right) .$$

Enfin, pour β_{33} , on a que

$$\begin{aligned} \beta_{33} &= 1 - \beta_{31} - \beta_{32} = 1 - \frac{3}{2N} + \frac{2}{4N^2} \\ &= \left(1 - \frac{1}{2N}\right) \left(1 - \frac{1}{N}\right) . \end{aligned}$$

On peut maintenant commencer le calcul exact de l'espérance de T_3 lorsqu'il n'y a pas de sélection. On a que

$$\begin{aligned} E_0(T_3) &= 1 \times \beta_{31} + 2 \times \left(\beta_{33} \beta_{31} + \beta_{32} \beta_{21}\right) + 3 \times \left(\beta_{33}^2 \beta_{31} + \beta_{33} \beta_{32} \beta_{21} + \beta_{32} \beta_{22} \beta_{21}\right) \\ &\quad + \dots + l \times \left(\beta_{33}^{l-1} \beta_{31} + \sum_{i=0}^{l-2} \left(\beta_{33}^i \beta_{32} \beta_{22}^{l-i-2} \beta_{21}\right)\right) + \dots , \\ &= \beta_{31} + \sum_{l=2}^{\infty} \left[l \beta_{33}^{l-1} \beta_{31}\right] + \sum_{l=2}^{\infty} \left[l \sum_{i=0}^{l-2} \left(\beta_{33}^i \beta_{32} \beta_{22}^{l-i-2} \beta_{21}\right)\right] , \\ &= \beta_{31} + \beta_{31} \left(\frac{1}{(1 - \beta_{33})^2} - 1\right) + \frac{\beta_{32} \beta_{21}}{\beta_{22}} \sum_{l=2}^{\infty} \left[l \beta_{22}^{l-1} \sum_{i=0}^{l-2} \left(\frac{\beta_{33}}{\beta_{22}}\right)^i\right] , \end{aligned}$$

car $0 < \beta_{33} < 1$. Puisque

$$\left(\frac{\beta_{33}}{\beta_{22}}\right)^i = \left(\frac{\left(1 - \frac{1}{2N}\right)\left(1 - \frac{1}{N}\right)}{1 - \frac{1}{2N}}\right)^i = \left(1 - \frac{1}{N}\right)^i < 1 ,$$

on a que

$$\sum_{i=0}^{l-2} \left(\frac{\beta_{33}}{\beta_{22}} \right)^i = \sum_{i=0}^{l-2} \left(1 - \frac{1}{N} \right)^i = \frac{1 - \left(1 - \frac{1}{N} \right)^{l-1}}{1 - \left(1 - \frac{1}{N} \right)} = N \left[1 - \left(1 - \frac{1}{N} \right)^{l-1} \right],$$

et il s'ensuit que

$$\begin{aligned} E_0(T_3) &= \beta_{31} + \beta_{31} \left(\frac{1}{(1 - \beta_{33})^2} - 1 \right) + \frac{\beta_{32} \beta_{21}}{\beta_{22}} \sum_{l=2}^{\infty} \left[l \beta_{22}^{l-1} N \left(1 - \left(1 - \frac{1}{N} \right)^{l-1} \right) \right], \\ &= \beta_{31} + \beta_{31} \left(\frac{1}{(1 - \beta_{33})^2} - 1 \right) \\ &\quad + \frac{N \beta_{32} \beta_{21}}{\beta_{22}} \left[\sum_{l=2}^{\infty} \left(l \beta_{22}^{l-1} \right) - \sum_{l=2}^{\infty} l \left(\beta_{22} \left(1 - \frac{1}{N} \right) \right)^{l-1} \right], \\ &= \beta_{31} + \beta_{31} \left(\frac{1}{(1 - \beta_{33})^2} - 1 \right) \\ &\quad + \frac{N \beta_{32} \beta_{21}}{\beta_{22}} \left[\sum_{l=1}^{\infty} \left(l \beta_{22}^{l-1} \right) - 1 - \sum_{l=1}^{\infty} l \left(\beta_{22} \left(1 - \frac{1}{N} \right) \right)^{l-1} + 1 \right], \\ &= \beta_{31} + \beta_{31} \left(\frac{1}{(1 - \beta_{33})^2} - 1 \right) + \frac{N \beta_{32} \beta_{21}}{\beta_{22}} \left[\frac{1}{(1 - \beta_{22})^2} - \frac{1}{\left[1 - \beta_{22} \left(1 - \frac{1}{N} \right) \right]^2} \right], \end{aligned}$$

car β_{22} et $\beta_{22}(1 - 1/N)$ sont entre 0 et 1. En remplaçant les β_{ij} par leur valeur respective, on obtient

$$\begin{aligned}
E_0(T_3) &= \frac{1}{4N^2} + \frac{1}{4N^2} \left(\frac{1}{\left(\frac{3}{2N} - \frac{2}{4N^2}\right)^2} - 1 \right) + \frac{\frac{3}{2}\left(1 - \frac{1}{2N}\right)\frac{1}{2N}}{1 - \frac{1}{2N}} \left(\frac{1}{\left(\frac{1}{2N}\right)^2} - \frac{1}{\left(\frac{3}{2N} - \frac{2}{4N^2}\right)^2} \right), \\
&= \frac{1}{4N^2} + \frac{1}{4N^2} \left(\left(\frac{4N^2}{6N-2} \right)^2 - 1 \right) + \frac{3}{4N} \left(4N^2 - \left(\frac{4N^2}{6N-2} \right)^2 \right), \\
&= \frac{1}{4N^2} + \frac{4N^2}{(6N-2)^2} - \frac{1}{4N^2} + 3N - \frac{12N^3}{(6N-2)^2}, \\
&= \frac{N(96N^2 - 68N + 12)}{(6N-2)^2}, \\
&= \frac{N(8N-3)}{3N-1}.
\end{aligned}$$

Annexe B

LES MATRICES

B.1. MATRICE H DANS LE CAS DE L'AUTOFÉCONDATION PARTIELLE

Soit Π la matrice de transition lorsqu'on suit les gènes ancestraux dans une population diploïde avec autofécondation partielle. On peut écrire $\Pi = H + R_N/2N$ et H contient les probabilités de transition (d'ordre 1) associées à l'autofécondation. Pour trouver les éléments de H , on trouve d'abord ceux de la matrice Π qui sont relativement simples à trouver. L'élément ij de H est constitué des termes de l'élément ij de Π qui ne dépendent pas de N . Voici les éléments de H qui est une matrice 15×15 :

$$H_{[1,1]} = 1, \quad H_{[1,j]} = 0, \text{ pour } j \geq 2,$$

$$H_{[2,1]} = 0, \quad H_{[2,2]} = 1, \quad H_{[2,j]} = 0, \text{ pour } j \geq 3,$$

$$H_{[3,1]} = \frac{\alpha}{2}, \quad H_{[3,2]} = 1 - \alpha, \quad H_{[3,3]} = \frac{\alpha}{2},$$

$$H_{[3,j]} = 0, \text{ pour } j \geq 4,$$

$$H_{[4,4]} = 1,$$

$$H_{[4,j]} = 0, \text{ pour } j \neq 4,$$

$$H_{[5,2]} = \frac{\alpha}{2},$$

$$H_{[5,4]} = 1 - \alpha,$$

$$H_{[5,5]} = \frac{\alpha}{2},$$

$$H_{[5,j]} = 0, \text{ pour } j \neq 2, 4, 5,$$

$$H_{[6,6]} = 1,$$

$$H_{[6,j]} = 0, \text{ pour } j \neq 6,$$

$$H_{[7,4]} = \frac{\alpha}{2},$$

$$H_{[7,6]} = 1 - \alpha,$$

$$H_{[7,7]} = \frac{\alpha}{2},$$

$$H_{[7,j]} = 0, \text{ pour } j \neq 4, 6, 7,$$

$$H_{[8,2]} = \frac{\alpha^2}{4},$$

$$H_{[8,4]} = (1 - \alpha)\alpha,$$

$$H_{[8,5]} = \frac{\alpha^2}{2},$$

$$H_{[8,6]} = (1 - \alpha)^2,$$

$$H_{[8,7]} = (1 - \alpha)\alpha,$$

$$H_{[8,8]} = \frac{\alpha^2}{4},$$

$$H_{[8,j]} = 0, \text{ pour } j \neq 2, 4, 5, 6, 7, 8,$$

$$H_{[9,9]} = 1,$$

$$H_{[9,j]} = 0, \text{ pour } j \neq 9,$$

$$H_{[10,6]} = \frac{\alpha}{2},$$

$$H_{[10,9]} = 1 - \alpha,$$

$$H_{[10,10]} = \frac{\alpha}{2},$$

$$H_{[10,j]} = 0, \text{ pour } j \neq 6, 9, 10,$$

$$H_{[11,4]} = \frac{\alpha^2}{4}, \quad H_{[11,6]} = (1 - \alpha)\alpha, \quad H_{[11,7]} = \frac{\alpha^2}{2},$$

$$H_{[11,9]} = (1 - \alpha)^2, \quad H_{[11,10]} = (1 - \alpha)\alpha, \quad H_{[11,11]} = \frac{\alpha^2}{4},$$

$$H_{[11,j]} = 0, \text{ pour } j \neq 4, 6, 7, 9, 10, 11,$$

$$H_{[12,12]} = 1, \quad H_{[12,j]} = 0, \text{ pour } j \neq 12,$$

$$H_{[13,9]} = \frac{\alpha}{2}, \quad H_{[13,12]} = 1 - \alpha, \quad H_{[13,13]} = \frac{\alpha}{2},$$

$$H_{[13,j]} = 0, \text{ pour } j \neq 9, 12, 13,$$

$$H_{[14,6]} = \frac{\alpha^2}{4}, \quad H_{[14,9]} = (1 - \alpha)\alpha, \quad H_{[14,10]} = \frac{\alpha^2}{2},$$

$$H_{[14,12]} = (1 - \alpha)^2, \quad H_{[14,13]} = (1 - \alpha)\alpha, \quad H_{[14,14]} = \frac{\alpha^2}{4},$$

$$H_{[14,j]} = 0, \text{ pour } j \neq 6, 9, 10, 12, 13, 14,$$

$$H_{[15,4]} = \frac{\alpha^3}{8}, \quad H_{[15,6]} = \frac{3\alpha^2(1 - \alpha)}{4}, \quad H_{[15,7]} = \frac{3\alpha^3}{8},$$

$$H_{[15,9]} = \frac{3\alpha(1 - \alpha)^2}{2}, \quad H_{[15,10]} = \frac{3\alpha^2(1 - \alpha)}{2}, \quad H_{[15,11]} = \frac{3\alpha^3}{8},$$

$$H_{[15,12]} = (1 - \alpha)^3,$$

$$H_{[15,13]} = \frac{3\alpha(1 - \alpha)^2}{2},$$

$$H_{[15,14]} = \frac{3\alpha^2(1 - \alpha)}{4},$$

$$H_{[15,15]} = \frac{\alpha^3}{8},$$

$$H_{[15,j]} = 0, \quad \text{pour } j = 1, 2, 3, 5, 8.$$

B.2. MATRICE P DANS LE CAS DE L'AUTOFÉCONDATION PARTIELLE

Pour utiliser le lemme de Möhle (lemme C.0.1), on doit trouver la matrice P telle que $P_{[i,j]} = \lim_{n \rightarrow \infty} (H^n)_{[i,j]}$. Puisque les éléments de la matrice H sont donnés dans l'annexe B.1, ceux de la matrice P peuvent donc être calculés. Voici les éléments de P qui est aussi une matrice 15×15 :

$$P_{[1,1]} = 1, \quad P_{[1,j]} = 0, \quad \text{pour } j \geq 2,$$

$$P_{[2,1]} = 0, \quad P_{[2,2]} = 1, \quad P_{[2,j]} = 0, \quad \text{pour } j \geq 3,$$

$$P_{[3,1]} = \frac{\alpha}{2 - \alpha}, \quad P_{[3,2]} = \frac{2(1 - \alpha)}{2 - \alpha}, \quad P_{[3,j]} = 0, \quad \text{pour } j \geq 3,$$

$$P_{[4,4]} = 1, \quad P_{[4,j]} = 0, \quad \text{pour } j \neq 4,$$

$$P_{[5,2]} = \frac{\alpha}{2 - \alpha}, \quad P_{[5,4]} = \frac{2(1 - \alpha)}{2 - \alpha}, \quad P_{[5,j]} = 0, \quad \text{pour } j \neq 2, 4,$$

$$P_{[6,6]} = 1, \quad P_{[6,j]} = 0, \quad \text{pour } j \neq 6,$$

$$P_{[7,4]} = \frac{\alpha}{2 - \alpha}, \quad P_{[7,6]} = \frac{2(1 - \alpha)}{2 - \alpha}, \quad P_{[7,j]} = 0, \quad \text{pour } j \neq 4, 6,$$

$$P_{[8,2]} = \left(\frac{\alpha}{2 - \alpha} \right)^2, \quad P_{[8,4]} = \frac{4(1 - \alpha)\alpha}{(2 - \alpha)^2}, \quad P_{[8,6]} = \left(\frac{2(1 - \alpha)}{2 - \alpha} \right)^2,$$

$$P_{[8,j]} = 0, \quad \text{pour } j \neq 2, 4, 6,$$

$$P_{[9,9]} = 1, \quad P_{[9,j]} = 0, \quad \text{pour } j \neq 9,$$

$$P_{[10,6]} = \frac{\alpha}{2-\alpha}, \quad P_{[10,9]} = \frac{2(1-\alpha)}{2-\alpha}, \quad P_{[10,j]} = 0, \quad \text{pour } j \neq 6, 9,$$

$$P_{[11,4]} = \left(\frac{\alpha}{2-\alpha} \right)^2, \quad P_{[11,6]} = \frac{4(1-\alpha)\alpha}{(2-\alpha)^2}, \quad P_{[11,9]} = \left(\frac{2(1-\alpha)}{2-\alpha} \right)^2,$$

$$P_{[11,j]} = 0, \quad \text{pour } j \neq 4, 6, 9,$$

$$P_{[12,12]} = 1, \quad P_{[12,j]} = 0, \quad \text{pour } j \neq 12,$$

$$P_{[13,9]} = \frac{\alpha}{2-\alpha}, \quad P_{[13,12]} = \frac{2(1-\alpha)}{2-\alpha}, \quad P_{[13,j]} = 0, \quad \text{pour } j \neq 9, 12,$$

$$P_{[14,6]} = \left(\frac{\alpha}{2-\alpha} \right)^2, \quad P_{[14,9]} = \frac{4(1-\alpha)\alpha}{(2-\alpha)^2}, \quad P_{[14,12]} = \left(\frac{2(1-\alpha)}{2-\alpha} \right)^2,$$

$$P_{[14,j]} = 0, \quad \text{pour } j \neq 6, 9, 12,$$

$$P_{[15,4]} = \left(\frac{\alpha}{2-\alpha} \right)^3, \quad P_{[15,6]} = \frac{6(1-\alpha)\alpha^2}{(2-\alpha)^3}, \quad P_{[15,9]} = \frac{12(1-\alpha)^2\alpha}{(2-\alpha)^3},$$

$$P_{[15,12]} = \left(\frac{2(1-\alpha)}{2-\alpha} \right)^3, \quad P_{[15,j]} = 0, \quad \text{pour } j \neq 4, 6, 9, 12.$$

B.3. MATRICE R_N DANS LE CAS DE L'AUTOFÉCONDATION PARTIELLE

La matrice R_N contient les probabilités de transition associées à la reproduction au hasard telle que $\Pi = H + R_N/2N$. Puisque c'est le produit $P(\lim_{N \rightarrow \infty} R_N)P$ qui nous intéresse, on peut ne trouver que les éléments des lignes 1, 2, 4, 6, 9, 12 de R_N , car les colonnes correspondantes de P sont les seules qui soient non nulles. Pour trouver les éléments de R_N , on doit d'abord trouver ceux de la matrice Π (ils sont relativement faciles à trouver). Puisque l'élément ij de Π est constitué de termes d'ordre $1/N^k$, $k \geq 0$, alors l'élément ij de R_N est constitué des termes d'ordre $1/N^k$, $k \geq 1$, de $\Pi_{[i,j]}$ qu'on multiplie par le facteur $2N$. Voici ces éléments de R_N qui est une matrice 15×15 :

$$R_{N[1,j]} = 0, \quad \text{pour tout } j,$$

$$R_{N[2,1]} = 1, \quad R_{N[2,2]} = -2, \quad R_{N[2,3]} = 1,$$

$$R_{N[2,j]} = 0, \quad \text{pour } j \geq 4,$$

$$R_{N[4,1]} = 1/2N, \quad R_{N[4,2]} = 3 - 3/N, \quad R_{N[4,3]} = 3/2N,$$

$$R_{N[4,4]} = -6 + 4/N, \quad R_{N[4,5]} = 3 - 3/N, \quad R_{N[4,j]} = 0, \quad \text{pour } j \geq 6,$$

$$R_{N[6,1]} = 1/(4N^2), \quad R_{N[6,2]} = \frac{7(-1 + N)}{2N^2}, \quad R_{N[6,3]} = 7/(4N^2),$$

$$R_{N[6,4]} = \frac{6(-2 + N)(-1 + N)}{N^2}, \quad R_{N[6,5]} = \frac{9(-1 + N)}{N^2},$$

$$R_{N[6,6]} = \frac{2(-6 + 11N - 6N^2)}{N^2},$$

$$R_{N[6,7]} = \frac{6(-2 + N)(-1 + N)}{N^2},$$

$$R_{N[6,8]} = \frac{3(-1 + N)}{2N^2}, \quad R_{N[6,j]} = 0, \quad \text{pour } j \geq 9,$$

$$R_{N[9,1]} = 1/(8N^3), \quad R_{N[9,2]} = \frac{15(-1 + N)}{4N^3}, \quad R_{N[9,3]} = 15/(8N^3),$$

$$R_{N[9,4]} = \frac{25(-2 + N)(-1 + N)}{2N^3},$$

$$R_{N[9,5]} = \frac{75(-1 + N)}{4N^3},$$

$$R_{N[9,6]} = \frac{10(-3 + N)(-2 + N)(-1 + N)}{N^3},$$

$$R_{N[9,7]} = \frac{30(-2 + N)(-1 + N)}{N^3},$$

$$R_{N[9,8]} = \frac{15(-1 + N)}{2N^3},$$

$$R_{N[9,9]} = \frac{48 - 100N + 70N^2 - 20N^3}{N^3},$$

$$R_{N[9,10]} = \frac{10(-3 + N)(-2 + N)(-1 + N)}{N^3},$$

$$R_{N[9,11]} = \frac{15(-2 + N)(-1 + N)}{2N^3},$$

$$R_{N[9,j]} = 0, \quad \text{pour } j \geq 12,$$

$$R_{N[12,1]} = 1/(16N^4),$$

$$R_{N[12,2]} = \frac{31(-1 + N)}{8N^4},$$

$$R_{N[12,3]} = 31/(16N^4),$$

$$R_{N[12,4]} = \frac{45(-2 + N)(-1 + N)}{2N^4},$$

$$R_{N[12,5]} = \frac{135(-1 + N)}{4N^4},$$

$$R_{N[12,6]} = \frac{65(-3 + N)(-2 + N)(-1 + N)}{2N^4},$$

$$R_{N[12,7]} = \frac{195(-2 + N)(-1 + N)}{2N^4},$$

$$R_{N[12,8]} = \frac{195(-1 + N)}{8N^4},$$

$$R_{N[12,9]} = \frac{15(-4+N)(-3+N)(-2+N)(-1+N)}{N^4},$$

$$R_{N[12,10]} = \frac{75(-3+N)(-2+N)(-1+N)}{N^4},$$

$$R_{N[12,11]} = \frac{225(-2+N)(-1+N)}{4N^4},$$

$$R_{N[12,12]} = \frac{-2(120 - 274N + 225N^2 - 85N^3 + 15N^4)}{N^4},$$

$$R_{N[12,13]} = \frac{15(-4+N)(-3+N)(-2+N)(-1+N)}{N^4},$$

$$R_{N[12,14]} = \frac{45(-3+N)(-2+N)(-1+N)}{2N^4},$$

$$R_{N[12,15]} = \frac{15(-2+N)(-1+N)}{4N^4}.$$

B.4. MATRICE R DANS LE CAS DE L'AUTOFÉCONDATION PARTIELLE

La matrice R est la matrice telle que $\lim_{N \rightarrow \infty} R_N = R$. La matrice R contient donc les probabilités limite de transition associées à la reproduction au hasard. L'élément ij de R est obtenu en prenant la limite, lorsque N tend vers l'infini, de l'élément ij de la matrice R_N . On a déjà indiqué, puisque c'est le produit PRP qui nous intéresse, qu'on peut ne trouver que les éléments des lignes 1, 2, 4, 6, 9, 12 de R , car les colonnes correspondantes de P sont les seules qui soient non nulles. Voici ces éléments de la matrice R qui est de taille 15×15 :

$$R_{[1,j]} = 0, \quad \text{pour tout } j,$$

$$R_{[2,1]} = 1, \quad R_{[2,2]} = -2, \quad R_{[2,3]} = 1,$$

$$R_{[2,j]} = 0, \quad \text{pour } j \geq 4,$$

$$R_{[4,2]} = 3, \quad R_{[4,4]} = -6, \quad R_{[4,5]} = 3,$$

$$R_{[4,j]} = 0, \quad \text{pour } j \neq 2, 4, 5,$$

$$R_{[6,4]} = 6, \quad R_{[6,6]} = -12, \quad R_{[6,7]} = 6,$$

$$R_{[6,j]} = 0, \quad \text{pour } j \neq 4, 6, 7,$$

$$R_{[9,6]} = 10, \quad R_{[9,9]} = -20, \quad R_{[9,10]} = 10,$$

$$R_{[9,j]} = 0, \quad \text{pour } j \neq 6, 9, 10,$$

$$R_{[12,9]} = 15,$$

$$R_{[12,12]} = -30,$$

$$R_{[12,13]} = 15,$$

$$R_{[12,j]} = 0, \quad \text{pour } j \neq 9, 12, 13.$$

B.5. MATRICE $e^{PRP\tau}$ DANS LE CAS DE L'AUTOFÉCONDATION PARTIELLE

Les termes de la matrice $e^{PRP\tau}$ doivent être trouvés pour utiliser le lemme de Möhle (C.0.1). Ayant déjà les éléments des matrices P et R (voir B.2 et B.4), on peut trouver la matrice G telle que $G = PRP$. En multipliant cette dernière par le scalaire τ et en prenant ensuite la matrice exponentielle, on obtient la matrice $e^{PRP\tau}$. Voici les éléments de celle-ci, qui est de taille 15×15 :

$$(e^{PRP\tau})_{[1,1]} = 1, \quad (e^{PRP\tau})_{[1,j]} = 0, \text{ pour } j \geq 2,$$

$$(e^{PRP\tau})_{[2,1]} = 1 - e^{-2\tau/(2-\alpha)}, \quad (e^{PRP\tau})_{[2,2]} = e^{-2\tau/(2-\alpha)},$$

$$(e^{PRP\tau})_{[2,j]} = 0, \text{ pour } j \geq 3,$$

$$(e^{PRP\tau})_{[3,1]} = \frac{2(1 - e^{-2\tau/(2-\alpha)})(1 - \alpha)}{2 - \alpha}, \quad (e^{PRP\tau})_{[3,2]} = -\frac{2(1 - e^{-2\tau/(2-\alpha)})(1 - \alpha)}{2 - \alpha},$$

$$(e^{PRP\tau})_{[3,3]} = 1, \quad (e^{PRP\tau})_{[3,j]} = 0, \text{ pour } j \geq 4,$$

$$(e^{PRP\tau})_{[4,1]} = \frac{1}{2} (1 - e^{-2\tau/(2-\alpha)})^2 (2 + e^{-2\tau/(2-\alpha)}),$$

$$(e^{PRP\tau})_{[4,2]} = \frac{3}{2} (e^{-2\tau/(2-\alpha)}) (1 - e^{-4\tau/(2-\alpha)}),$$

$$(e^{PRP\tau})_{[4,4]} = e^{-6\tau/(2-\alpha)},$$

$$(e^{PRP\tau})_{[4,j]} = 0, \text{ pour } j \neq 1, 2, 4,$$

$$(e^{PRP\tau})_{[5,1]} = \frac{1}{2-\alpha} \left(e^{-2\tau/(2-\alpha)}(2\alpha-3) + e^{-6\tau/(2-\alpha)}(1-\alpha) - \alpha + 2 \right),$$

$$(e^{PRP\tau})_{[5,2]} = \frac{1}{2-\alpha} \left(e^{-2\tau/(2-\alpha)}(3-2\alpha) + 3e^{-6\tau/(2-\alpha)}(\alpha-1) - \alpha \right),$$

$$(e^{PRP\tau})_{[5,4]} = \frac{2}{2-\alpha} (1 - e^{-6\tau/(2-\alpha)})(\alpha-1),$$

$$(e^{PRP\tau})_{[5,5]} = 1,$$

$$(e^{PRP\tau})_{[5,j]} = 0, \quad \text{pour } j \neq 1, 2, 4, 5,$$

$$(e^{PRP\tau})_{[6,1]} = 1 - \frac{9}{5}(e^{-2\tau/(2-\alpha)}) + e^{-6\tau/(2-\alpha)} - \frac{1}{5}(e^{-12\tau/(2-\alpha)}),$$

$$(e^{PRP\tau})_{[6,2]} = \frac{3}{5}(e^{-2\tau/(2-\alpha)})(3 - 5 e^{-4\tau/(2-\alpha)} + 2 e^{-10\tau/(2-\alpha)}),$$

$$(e^{PRP\tau})_{[6,4]} = 2 (e^{-6\tau/(2-\alpha)})(1 - e^{-6\tau/(2-\alpha)}),$$

$$(e^{PRP\tau})_{[6,6]} = e^{-12\tau/(2-\alpha)},$$

$$(e^{PRP\tau})_{[6,j]} = 0, \quad \text{pour } j \neq 1, 2, 4, 6,$$

$$(e^{PRP\tau})_{[7,1]} = \frac{1}{10(2-\alpha)} \left(e^{-2\tau/(2-\alpha)}(21\alpha-36) + 5 e^{-6\tau/(2-\alpha)}(4-3\alpha) \right. \\ \left. + 4 e^{-12\tau/(2-\alpha)}(\alpha-1) + 10(2-\alpha) \right),$$

$$(e^{PRP\tau})_{[7,2]} = \frac{3 e^{-2\tau/(2-\alpha)}}{10(2-\alpha)} \left(5 e^{-4\tau/(2-\alpha)}(3\alpha-4) + 8 e^{-10\tau/(2-\alpha)}(1-\alpha) - 7\alpha + 12 \right),$$

$$(e^{PRP\tau})_{[7,4]} = \frac{1}{2-\alpha} \left(-1 + e^{-6\tau/(2-\alpha)} \right) \left(4 e^{-6\tau/(2-\alpha)} (\alpha - 1) + \alpha \right),$$

$$(e^{PRP\tau})_{[7,6]} = \frac{2}{2-\alpha} \left(1 - e^{-12\tau/(2-\alpha)} \right) (\alpha - 1),$$

$$(e^{PRP\tau})_{[7,7]} = 1,$$

$$(e^{PRP\tau})_{[7,j]} = 0, \text{ pour } j \neq 1, 2, 4, 6, 7,$$

$$(e^{PRP\tau})_{[8,1]} = -\frac{e^{-2\tau/(2-\alpha)}(11\alpha^2 - 42\alpha + 36)}{5(2-\alpha)^2} + \frac{2 e^{-6\tau/(2-\alpha)}(1-\alpha)}{2-\alpha} \\ - \frac{4 e^{-12\tau/(2-\alpha)}(\alpha-1)^2}{5(2-\alpha)^2} + 1,$$

$$(e^{PRP\tau})_{[8,2]} = \frac{1}{5(2-\alpha)^2} \left(e^{-2\tau/(2-\alpha)}(11\alpha^2 - 42\alpha + 36) - 30 e^{-6\tau/(2-\alpha)}(\alpha-2)(\alpha-1) \right. \\ \left. + 24 e^{-12\tau/(2-\alpha)}(\alpha-1)^2 - 5\alpha^2 \right),$$

$$(e^{PRP\tau})_{[8,4]} = \frac{4}{(2-\alpha)^2} \left(1 - e^{-6\tau/(2-\alpha)} \right) (\alpha - 1) \left(2 e^{-6\tau/(2-\alpha)} (\alpha - 1) + \alpha \right),$$

$$(e^{PRP\tau})_{[8,6]} = \frac{4}{(2-\alpha)^2} \left(-1 + e^{-12\tau/(2-\alpha)} \right) (\alpha - 1)^2,$$

$$(e^{PRP\tau})_{[8,8]} = 1,$$

$$(e^{PRP\tau})_{[8,j]} = 0, \text{ pour } j \neq 1, 2, 4, 6, 8,$$

$$(e^{PRP\tau})_{[9,1]} = \frac{1}{14} \left(-28 e^{-2\tau/(2-\alpha)} + 20 e^{-6\tau/(2-\alpha)} - 7 e^{-12\tau/(2-\alpha)} + e^{-20\tau/(2-\alpha)} + 14 \right),$$

$$(e^{PRP\tau})_{[9,2]} = 2 e^{-2\tau/(2-\alpha)} - \frac{30}{7} e^{-6\tau/(2-\alpha)} + 3 e^{-12\tau/(2-\alpha)} - \frac{5}{7} e^{-20\tau/(2-\alpha)},$$

$$(e^{PRP\tau})_{[9,4]} = \frac{5}{7} (e^{-6\tau/(2-\alpha)}) (-7 e^{-6\tau/(2-\alpha)} + 3 e^{-14\tau/(2-\alpha)} + 4),$$

$$(e^{PRP\tau})_{[9,6]} = \frac{5}{2} (e^{-12\tau/(2-\alpha)}) (1 - e^{-8\tau/(2-\alpha)}),$$

$$(e^{PRP\tau})_{[9,9]} = e^{-20\tau/(2-\alpha)},$$

$$(e^{PRP\tau})_{[9,j]} = 0, \text{ pour } j \neq 1, 2, 4, 6, 9,$$

$$(e^{PRP\tau})_{[10,1]} = \frac{1}{35} \left[\frac{7 e^{-2\tau/(2-\alpha)} (11\alpha - 20)}{2 - \alpha} + \frac{5 e^{-6\tau/(2-\alpha)} (20 - 13\alpha)}{2 - \alpha} \right. \\ \left. + \frac{7 e^{-12\tau/(2-\alpha)} (4\alpha - 5)}{2 - \alpha} + \frac{5 e^{-20\tau/(2-\alpha)} (1 - \alpha)}{2 - \alpha} + 35 \right],$$

$$(e^{PRP\tau})_{[10,2]} = \frac{e^{-2\tau/(2-\alpha)}}{35(2-\alpha)} \left(e^{-4\tau/(2-\alpha)} (195\alpha - 300) + 42 e^{-10\tau/(2-\alpha)} (5 - 4\alpha) \right. \\ \left. + 50 e^{-18\tau/(2-\alpha)} (\alpha - 1) - 77\alpha + 140 \right),$$

$$(e^{PRP\tau})_{[10,4]} = \frac{2 e^{-6\tau/(2-\alpha)}}{7(2-\alpha)} \left(7 e^{-6\tau/(2-\alpha)} (4\alpha - 5) + 15 e^{-14\tau/(2-\alpha)} (1 - \alpha) - 13\alpha + 20 \right),$$

$$(e^{PRP\tau})_{[10,6]} = \frac{1}{(2-\alpha)} \left(e^{-12\tau/(2-\alpha)} (5 - 4\alpha) + 5 e^{-20\tau/(2-\alpha)} (\alpha - 1) - \alpha \right),$$

$$(e^{PRP\tau})_{[10,9]} = \frac{2}{2-\alpha} (1 - e^{-20\tau/(2-\alpha)})(\alpha - 1),$$

$$(e^{PRP\tau})_{[10,10]} = 1,$$

$$(e^{PRP\tau})_{[10,j]} = 0, \text{ pour } j \neq 1, 2, 4, 6, 9, 10,$$

$$(e^{PRP\tau})_{[11,1]} = \frac{1}{70} \left[\frac{-7 e^{-2\tau/(2-\alpha)}(23\alpha^2 - 88\alpha + 80)}{(2-\alpha)^2} + \frac{5 e^{-6\tau/(2-\alpha)}(31\alpha^2 - 104\alpha + 80)}{(2-\alpha)^2} \right. \\ \left. - \frac{28 e^{-12\tau/(2-\alpha)}(3\alpha^2 - 8\alpha + 5)}{(2-\alpha)^2} + \frac{20 e^{-20\tau/(2-\alpha)}(\alpha - 1)^2}{(2-\alpha)^2} + 70 \right],$$

$$(e^{PRP\tau})_{[11,2]} \\ = \frac{e^{-2\tau/(2-\alpha)}}{70(2-\alpha)^2} \left[-15 e^{-4\tau/(2-\alpha)}(31\alpha^2 - 104\alpha + 80) + 168 e^{-10\tau/(2-\alpha)}(3\alpha^2 - 8\alpha + 5) \right. \\ \left. - 200 e^{-18\tau/(2-\alpha)}(\alpha - 1)^2 + 7(23\alpha^2 - 88\alpha + 80) \right],$$

$$(e^{PRP\tau})_{[11,4]} \\ = \frac{1}{7(2-\alpha)^2} \left[e^{-6\tau/(2-\alpha)}(31\alpha^2 - 104\alpha + 80) - 28 e^{-12\tau/(2-\alpha)}(3\alpha^2 - 8\alpha + 5) \right. \\ \left. + 60 e^{-20\tau/(2-\alpha)}(\alpha - 1)^2 - 7\alpha^2 \right],$$

$$(e^{PRP\tau})_{[11,6]} = \frac{2}{(2-\alpha)^2} \left(e^{-12\tau/(2-\alpha)}(3\alpha^2 - 8\alpha + 5) - 5 e^{-20\tau/(2-\alpha)}(\alpha - 1)^2 + 2\alpha(\alpha - 1) \right),$$

$$(e^{PRP\tau})_{[11,9]} = \frac{4}{(2-\alpha)^2} (-1 + e^{-20\tau/(2-\alpha)})(\alpha - 1)^2,$$

$$(e^{PRP\tau})_{[11,11]} = 1,$$

$$(e^{PRP\tau})_{[11,j]} = 0, \quad \text{pour } j \neq 1, 2, 4, 6, 9, 11,$$

$$(e^{PRP\tau})_{[12,1]} = \frac{1}{42} \left(-90 e^{-2\tau/(2-\alpha)} + 75 e^{-6\tau/(2-\alpha)} - 35 e^{-12\tau/(2-\alpha)} + 9 e^{-20\tau/(2-\alpha)} \right. \\ \left. - e^{-30\tau/(2-\alpha)} + 42 \right),$$

$$(e^{PRP\tau})_{[12,2]} = \frac{5}{14} \left(6 e^{-2\tau/(2-\alpha)} - 15 e^{-6\tau/(2-\alpha)} + 14 e^{-12\tau/(2-\alpha)} - 6 e^{-20\tau/(2-\alpha)} \right. \\ \left. + e^{-30\tau/(2-\alpha)} \right),$$

$$(e^{PRP\tau})_{[12,4]} = \frac{5 e^{-6\tau/(2-\alpha)}}{21} \left(-35 e^{-6\tau/(2-\alpha)} + 27 e^{-14\tau/(2-\alpha)} - 7 e^{-24\tau/(2-\alpha)} + 15 \right),$$

$$(e^{PRP\tau})_{[12,6]} = \frac{5 e^{-12\tau/(2-\alpha)}}{6} \left(-9 e^{-8\tau/(2-\alpha)} + 4 e^{-18\tau/(2-\alpha)} + 5 \right),$$

$$(e^{PRP\tau})_{[12,9]} = 3 e^{-20\tau/(2-\alpha)} (1 - e^{-10\tau/(2-\alpha)}),$$

$$(e^{PRP\tau})_{[12,12]} = e^{-30\tau/(2-\alpha)},$$

$$(e^{PRP\tau})_{[12,j]} = 0, \quad \text{pour } j \neq 1, 2, 4, 6, 9, 12,$$

$$(e^{PRP\tau})_{[13,1]} = \frac{1}{42(2-\alpha)} \left[e^{-2\tau/(2-\alpha)} (96\alpha - 180) + 30 e^{-6\tau/(2-\alpha)} (5 - 3\alpha) \right. \\ \left. + e^{-12\tau/(2-\alpha)} (49\alpha - 70) + 3 e^{-20\tau/(2-\alpha)} (6 - 5\alpha) \right. \\ \left. - 2 e^{-30\tau/(2-\alpha)} (1 - \alpha) + 42(2 - \alpha) \right],$$

$$(e^{PRP\tau})_{[13,2]} = \frac{e^{-2\tau/(2-\alpha)}}{7(2-\alpha)} \left[e^{-4\tau/(2-\alpha)}(45\alpha - 75) + 7 e^{-10\tau/(2-\alpha)}(10 - 7\alpha) \right. \\ \left. - 5 e^{-18\tau/(2-\alpha)}(6 - 5\alpha) + 5 e^{-28\tau/(2-\alpha)}(1 - \alpha) - 16\alpha + 30 \right],$$

$$(e^{PRP\tau})_{[13,4]} = \frac{5 e^{-6\tau/(2-\alpha)}}{21(2-\alpha)} \left[7 e^{-6\tau/(2-\alpha)}(7\alpha - 10) - 9 e^{-14\tau/(2-\alpha)}(5\alpha - 6) \right. \\ \left. + 14 e^{-24\tau/(2-\alpha)}(\alpha - 1) - 18\alpha + 30 \right],$$

$$(e^{PRP\tau})_{[13,6]} = \frac{5 e^{-12\tau/(2-\alpha)}}{6(2-\alpha)} \left[3 e^{-8\tau/(2-\alpha)}(5\alpha - 6) + 8 e^{-18\tau/(2-\alpha)}(1 - \alpha) - 7\alpha + 10 \right],$$

$$(e^{PRP\tau})_{[13,9]} = \frac{1}{(2-\alpha)} \left[e^{-20\tau/(2-\alpha)}(6 - 5\alpha) - 6 e^{-30\tau/(2-\alpha)}(1 - \alpha) - \alpha \right],$$

$$(e^{PRP\tau})_{[13,12]} = \frac{2}{(2-\alpha)} (1 - e^{-30\tau/(2-\alpha)})(\alpha - 1),$$

$$(e^{PRP\tau})_{[13,13]} = 1,$$

$$(e^{PRP\tau})_{[13,j]} = 0, \quad \text{pour } j \neq 1, 2, 4, 6, 9, 12, 13,$$

$$(e^{PRP\tau})_{[14,1]} = \frac{1}{105(2-\alpha)^2} \left[-3 e^{-2\tau/(2-\alpha)}(83\alpha^2 - 320\alpha + 300) + 15 e^{-6\tau/(2-\alpha)}(17\alpha^2 - 60\alpha + 50) \right. \\ \left. - 7 e^{-12\tau/(2-\alpha)}(23\alpha^2 - 70\alpha + 50) + 30 e^{-20\tau/(2-\alpha)}(2\alpha^2 - 5\alpha + 3) \right. \\ \left. - 10 e^{-30\tau/(2-\alpha)}(1 - \alpha)^2 + 105(2 - \alpha)^2 \right],$$

$$\begin{aligned}
(e^{PRP\tau})_{[14,2]} = & \\
& \frac{e^{-2\tau/(2-\alpha)}}{35(2-\alpha)^2} \left[-15 e^{-4\tau/(2-\alpha)}(17\alpha^2 - 60\alpha + 50) + 14 e^{-10\tau/(2-\alpha)}(23\alpha^2 - 70\alpha + 50) \right. \\
& \quad - 100 e^{-18\tau/(2-\alpha)}(2\alpha^2 - 5\alpha + 3) + 50 e^{-28\tau/(2-\alpha)}(1-\alpha)^2 \\
& \quad \left. + 83\alpha^2 - 320\alpha + 300 \right],
\end{aligned}$$

$$\begin{aligned}
(e^{PRP\tau})_{[14,4]} = & \\
& \frac{-2e^{-6\tau/(2-\alpha)}}{21(2-\alpha)^2} \left[7 e^{-6\tau/(2-\alpha)}(23\alpha^2 - 70\alpha + 50) - 90 e^{-14\tau/(2-\alpha)}(2\alpha^2 - 5\alpha + 3) \right. \\
& \quad \left. + 70 e^{-24\tau/(2-\alpha)}(1-\alpha)^2 - 3(17\alpha^2 - 60\alpha + 50) \right],
\end{aligned}$$

$$\begin{aligned}
(e^{PRP\tau})_{[14,6]} = & \frac{1}{3(2-\alpha)^2} \left[e^{-12\tau/(2-\alpha)}(23\alpha^2 - 70\alpha + 50) - 30 e^{-20\tau/(2-\alpha)}(2\alpha^2 - 5\alpha + 3) \right. \\
& \quad \left. + 40 e^{-30\tau/(2-\alpha)}(1-\alpha)^2 - 3\alpha^2 \right],
\end{aligned}$$

$$(e^{PRP\tau})_{[14,9]} = \frac{4}{(2-\alpha)^2} (1 - e^{-10\tau/(2-\alpha)})(\alpha - 1)(e^{-10\tau/(2-\alpha)}(\alpha) + 3 e^{-20\tau/(2-\alpha)}(\alpha - 1) + \alpha),$$

$$(e^{PRP\tau})_{[14,12]} = \frac{4}{(2-\alpha)^2} (-1 + e^{-30\tau/(2-\alpha)})(\alpha - 1)^2,$$

$$(e^{PRP\tau})_{[14,14]} = 1,$$

$$(e^{PRP\tau})_{[14,j]} = 0, \quad \text{pour } j \neq 1, 2, 4, 6, 9, 12, 14,$$

$$(e^{PRP\tau})_{[15,1]} = \frac{1}{210} \left[\frac{9 e^{-2\tau/(2-\alpha)} (57\alpha^3 - 332\alpha^2 + 640\alpha - 400)}{(2-\alpha)^3} + \frac{15 e^{-6\tau/(2-\alpha)} (37\alpha^2 - 130\alpha + 100)}{(2-\alpha)^2} \right. \\ \left. + \frac{28 e^{-12\tau/(2-\alpha)} (14\alpha^3 - 69\alpha^2 + 105\alpha - 50)}{(2-\alpha)^3} + \frac{180 e^{-20\tau/(2-\alpha)} (1-\alpha)^2}{(2-\alpha)^2} \right. \\ \left. + \frac{40 e^{-30\tau/(2-\alpha)} (\alpha-1)^3}{(2-\alpha)^3} + 210 \right],$$

$$(e^{PRP\tau})_{[15,2]} = \frac{e^{-2\tau/(2-\alpha)}}{70(2-\alpha)^3} \left[15 e^{-4\tau/(2-\alpha)} (37\alpha^3 - 204\alpha^2 + 360\alpha - 200) \right. \\ \left. - 56 e^{-10\tau/(2-\alpha)} (14\alpha^3 - 69\alpha^2 + 105\alpha - 50) \right. \\ \left. + 600 e^{-18\tau/(2-\alpha)} (\alpha-2)(1-\alpha)^2 \right. \\ \left. - 200 e^{-28\tau/(2-\alpha)} (\alpha-1)^3 - 3(57\alpha^3 - 332\alpha^2 + 640\alpha - 400) \right],$$

$$(e^{PRP\tau})_{[15,4]} = \frac{1}{21(2-\alpha)^3} \left[-3 e^{-6\tau/(2-\alpha)} (37\alpha^3 - 204\alpha^2 + 360\alpha - 200) \right. \\ \left. + 28 e^{-12\tau/(2-\alpha)} (14\alpha^3 - 69\alpha^2 + 105\alpha - 50) \right. \\ \left. - 540 e^{-20\tau/(2-\alpha)} (\alpha-2)(1-\alpha)^2 \right. \\ \left. + 280 e^{-30\tau/(2-\alpha)} (\alpha-1)^3 - 21\alpha^3 \right],$$

$$(e^{PRP\tau})_{[15,6]} = \frac{2}{3(2-\alpha)^3} \left[-e^{-12\tau/(2-\alpha)} (14\alpha^3 - 69\alpha^2 + 105\alpha - 50) \right. \\ \left. + 45 e^{-20\tau/(2-\alpha)} (\alpha-2)(1-\alpha)^2 \right. \\ \left. - 40 e^{-30\tau/(2-\alpha)} (\alpha-1)^3 + 9\alpha^2(\alpha-1) \right],$$

$$(e^{PRP\tau})_{[15,9]} = \frac{12}{(2-\alpha)^3} \left[(-1 + e^{-10\tau/(2-\alpha)})(1-\alpha)^2 \right. \\ \left. \times (e^{-10\tau/(2-\alpha)}(\alpha) + 2 e^{-20\tau/(2-\alpha)}(\alpha-1) + \alpha) \right],$$

$$(e^{PRP\tau})_{[15,12]} = \frac{8}{(2-\alpha)^3} (1 - e^{-30\tau/(2-\alpha)}) (\alpha - 1)^3,$$

$$(e^{PRP\tau})_{[15,15]} = 1,$$

$$(e^{PRP\tau})_{[15,j]} = 0, \quad \text{pour } j \neq 1, 2, 4, 6, 9, 12, 15.$$

Annexe C

LEMME DE MÖHLE ET AUTOFÉCONDATION PARTIELLE

Soient $H_{[i,j]}$ les éléments de la matrice H . On définit la norme

$$\|H\| := \max_i \sum_j |H_{[i,j]}|.$$

Soit $[\tau]$ le plus grand entier plus petit que τ .

Lemme C.0.1. *Lemme de Möhle : (voir [M])*

Soient des constantes $\tau, K \geq 0$ et soit $\{c_N\}_{N \in \mathbb{N}}$ une suite de nombres réels positifs telle que $\lim_{N \rightarrow \infty} c_N = 0$. Soit H une matrice telle que $\|H\| = 1$ et que $P := \lim_{n \rightarrow \infty} H^n$ existe. Alors

$$\lim_{N \rightarrow \infty} \sup_{\|R\| \leq K} \|(H + c_N R)^{[\tau/c_N]} - (P + c_N R)^{[\tau/c_N]}\| = 0.$$

Si $\{R_N\}_{N \in \mathbb{N}}$ est une suite de matrices telle que $G := \lim_{N \rightarrow \infty} P R_N P$ existe, alors

$$\lim_{N \rightarrow \infty} (H + c_N R_N)^{[\tau/c_N]} = P - I + e^{\tau G} = P e^{\tau G} \quad \forall \tau \geq 0.$$

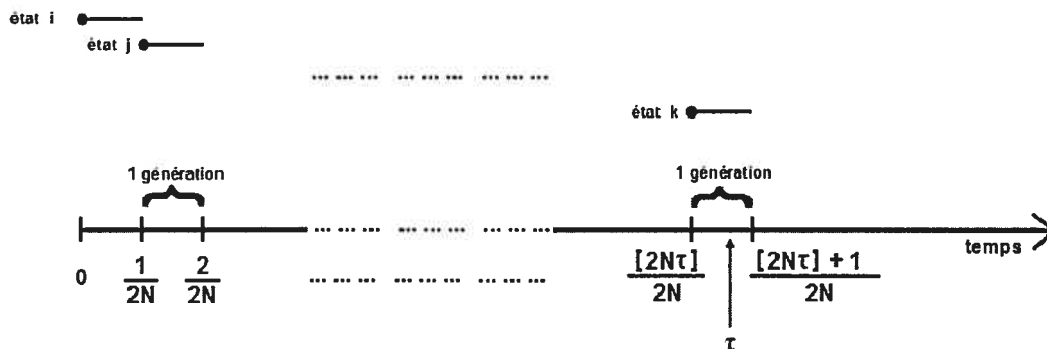
Il est à noter que dans les cas où $R := \lim_{N \rightarrow \infty} R_N$ existe, alors $G = PRP$. Expliquons la façon dont on se sert de la deuxième partie de ce lemme.

Soit Π la matrice de transition sur l'ensemble de toutes les configurations possibles de gènes ancestraux d'une génération à la précédente incluant l'éventualité de coalescence. En faisant ainsi, le nombre de gènes ancestraux ne peut augmenter, il ne fait que diminuer par des coalescences. On veut $\sum_{t \geq 0} \Pi^t \times \frac{1}{2N}$, où t est une variable représentant les générations discrètes. Puisque Π est souvent une matrice de taille considérable, on cherche à simplifier le calcul. D'abord, on passe d'un comptage par nombre de générations à un comptage par nombre de $2N$ générations, c'est-à-dire que le nouveau temps τ correspondra à $2N\tau$ générations. La matrice de transition de 0 à τ dans le passé est alors $\Pi^{2N\tau}$ si $2N\tau$ est un entier. Si $2N\tau$ n'est pas un entier, on prend comme matrice de transition $\Pi^{[2N\tau]}$, où $[2N\tau]$ est la valeur entière de $2N\tau$ (voir la figure C.1). le lemme de Möhle est utilisé avec $c_N = 1/2N$ et Π sous forme $H + c_N R_N = A + \frac{R_N}{2N}$. On aura alors

$$\lim_{N \rightarrow \infty} \Pi^{[2N\tau]} = \lim_{N \rightarrow \infty} \left(H + \frac{R_N}{2N} \right)^{[2N\tau]} = P e^{\tau G}.$$

Comme la somme sur $t \geq 0$ de $(\Pi^t)_{[i,j]} \times \frac{1}{2N}$ peut être représentée par celle d'aires de rectangles dont la base est $1/2N$ et la hauteur est $(\Pi^t)_{[i,j]}$, cette somme sera alors approximée par l'intégrale de $[P e^{\tau G}]_{[i,j]}$ sur $\tau \geq 0$ lorsque N tend vers l'infini.

Voici un exemple tiré de [M] mais un peu modifié. On considère le modèle de Wright-Fisher appliqué à une population de N individus diploïdes ($2N$ gènes).

FIG. C.1. $\Pi^{[2N\tau]}$

On suppose qu'il y a autofécondation partielle, c'est-à-dire qu'un individu se reproduit avec lui-même avec probabilité α et avec un autre individu au hasard avec probabilité $1 - \alpha$. On s'intéresse à un seul locus pour lequel il y a deux allèles : A et B . On prend un échantillon de deux gènes parmi les $2N$ à partir desquels nous suivrons les gènes ancestraux. Avec deux lignées au maximum dans une population diploïde, il y a trois configurations possibles de gènes ancestraux (les gènes ancestraux sont en noir) :



FIG. C.2. Les trois états possibles pour la configuration ancestrale de deux gènes.

Dans la première configuration, il n'y a qu'un gène ancestral restant ; dans la deuxième, les deux gènes ancestraux sont dans deux individus différents ; dans la troisième, les deux gènes ancestraux sont dans le même individu. Ces configurations sont les trois états de la matrice Π dans l'ordre. Voici donc la matrice Π :

$$\Pi = \begin{pmatrix} 1 & 0 & 0 \\ 1/2N & 1 - 1/N & 1/2N \\ \alpha/2 & 1 - \alpha & \alpha/2 \end{pmatrix}.$$

On voit que Π peut s'écrire sous forme $H + c_N R_N = H + \frac{R_N}{2N}$, où

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha/2 & 1 - \alpha & \alpha/2 \end{pmatrix}, \quad R_N = R = \begin{pmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

La matrice H contient les probabilités de transition associées à l'autofécondation, alors que la matrice R_N contient les probabilités de transition associées à la reproduction au hasard. Ces deux matrices représentent deux échelles de temps différentes car le mécanisme associé à la matrice R est d'ordre N et donc plus lent que celui associé à la matrice H . On a

$$\begin{aligned} P &= \lim_{n \rightarrow \infty} H^n = \lim_{n \rightarrow \infty} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{\alpha}{2} \frac{1 - (\alpha/2)^n}{1 - \alpha/2} & (1 - \alpha) \frac{1 - (\alpha/2)^n}{1 - \alpha/2} & (\frac{\alpha}{2})^n \end{pmatrix}, \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{\alpha}{2 - \alpha} & \frac{2(1 - \alpha)}{2 - \alpha} & 0 \end{pmatrix}. \end{aligned}$$

La matrice P représente la matrice de transition "instantanée" lorsque N tend vers l'infini. On a vu en effet que le mécanisme associé à la matrice R est plus lent que celui associé à la matrice H et donc les changements d'état attribués au

mécanisme de H paraissent instantanés à la limite, lorsque N tend vers l'infini. Par ce fait, les états dans lesquels il y a deux gènes ancestraux dans le même individu (l'état 3 ici) paraissent instantanés, car très rapidement pour ces états, en remontant le temps, soit les lignées dans le même individu vont coalescer, soit ils vont se séparer dans deux individus différents. On peut maintenant appliquer le lemme C.0.1 :

$$\lim_{N \rightarrow \infty} \Pi^{[2N\tau]} = \lim_{N \rightarrow \infty} (H + R/2N)^{[2N\tau]} = P e^{\tau PRP},$$

où la matrice P , devant $e^{\tau PRP}$, représente la première transition instantanée au temps τ alors que la matrice $e^{\tau PRP}$ représente le processus plus lent, jusqu'au temps 0. On obtient que

$$\lim_{N \rightarrow \infty} \Pi^{[2N\tau]} = \begin{pmatrix} 1 & 0 & 0 \\ 1 - e^{-2t/(2-\alpha)} & e^{-2t/(2-\alpha)} & 0 \\ 1 - \frac{2(1-\alpha)e^{-2t/(2-\alpha)}}{2-\alpha} & \frac{2(1-\alpha)e^{-2t/(2-\alpha)}}{2-\alpha} & 0 \end{pmatrix}.$$